(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 December 2001 (13.12.2001)

PCT

(10) International Publication Number
WO 01/94546 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US01/17928

(22) International Filing Date: 31 May 2001 (31.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/585,768 2 June 2000 (02.06.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/585,768 (CIP)
Filed on 2 June 2000 (02.06.2000)

(71) Applicant *(for all designated States except US)*: DNA SCIENCES, INC. [US/US]; 6540 Kaiser Drive, Fremont, CA 94555-3613 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: XU, Hua [CN/US]; 1567 Samedra St, Sunnyvale, CA 94087 (US).

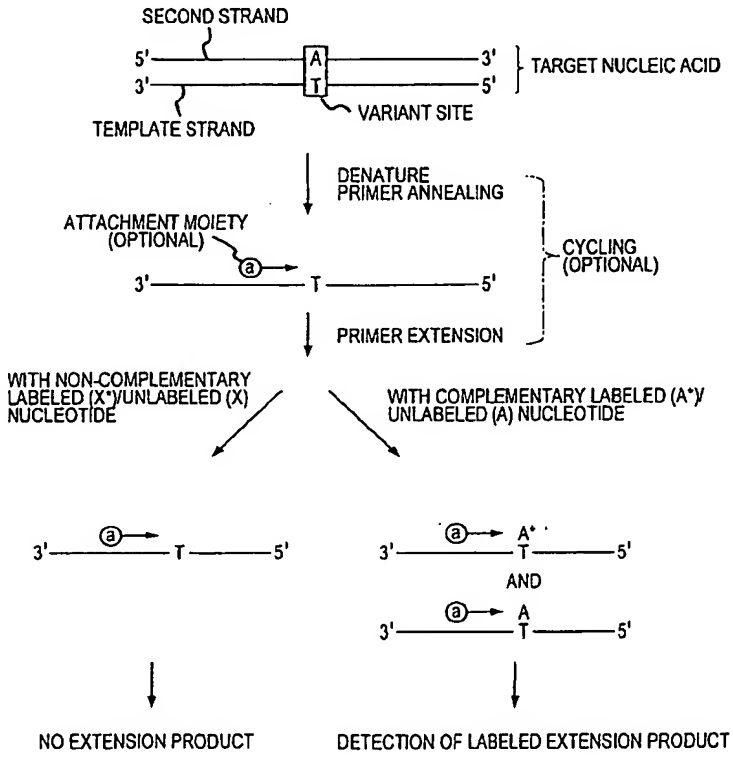
(74) Agents: AUSENHUS, Scott, L. et al.; Townsend and Townsend and Crew LLP, 1200 Seventeenth Street, Ste 2700, Denver, CO 80202 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

[Continued on next page]

(54) Title: PRIMER EXTENSION USING A MIXTURE OF LABELLED AND UNLABELLED NUCLEOTIDES



(57) Abstract: The present invention provides methods, compositions and kits for determining the identity of a nucleotide at a variant site in a nucleic acid of interest, including point mutations and single nucleotide polymorphisms. The methods utilize one or more nucleotides, each nucleotide being a mixture of labelled and unlabelled forms, to generate labelled extension products that are characteristic of the nucleotide at the variant site in the nucleic acid of interest. In addition to their utility in detecting and analyzing point mutations and SNPs, the methods and kits of the invention have utility in a variety of other applications in which specific nucleic acid sequence information is of value, including detection of pathogens, paternity disputes, prenatal testing and forensic analysis.

WO 01/94546 A2

patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

PRIMER EXTENSION USING A MIXTURE OF LABELLED AND UNLABELLED NUCLEOTIDES

CROSS-REFERENCES TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Application No. 09/585,768, filed June 2, 2000, which is incorporated herein in its entirety for all purposes.

FIELD OF INVENTION

The present invention relates to the field of molecular genetics, particularly the identification and detection of certain nucleotide sequences.

BACKGROUND OF THE INVENTION

The nucleic acids comprising the genome of an organism contain the genetic information for that organism. The translation or expression of these nucleic acids generates proteins that function in many diverse ways within the organism. Even minute changes in a nucleotide sequence, including single base pair substitutions, can have a significant effect in the quality or quantity of a protein. Single nucleotide changes are referred to as single nucleotide polymorphisms or simply SNPs, and the site at which the SNP occurs is typically referred to as a polymorphic site.

Many SNPs, as well as larger nucleic acid alterations, can affect the phenotype of the organism, and in some instances can result in the onset of disease. For example, diseases associated with SNPs include: sickle cell anemia, β-thalassemias, diabetes, cystic fibrosis, hyperlipoproteinemia, a wide variety of autoimmune diseases, and the formation of oncogenes. In addition to causing or affecting disease states, point mutations can cause altered pathogenicity and resistance to therapeutics that target certain microorganisms.

The ability to detect specific nucleotide alterations or mutations in DNA sequences has a number of medical and non-medical utilities. For example, methods capable of identifying nucleotide alterations provide a means for screening and diagnosing many common diseases that are associated with SNPs. Methods that can quickly identify such changes or mutations are also valuable in taking prophylactic measures, assessing the propensity for disease, and in patient counseling and education. As for non-medical

applications, such methods have value in the detection of microorganisms, resolving paternity disputes and in forensic analysis to identify perpetrators of crimes.

Various methods have been developed to obtain sequence information for variants sites. Such methods include hybridization reactions between a target nucleic acid and allele-specific oligonucleotide (ASO) probes (see, *e.g.*, European Patent Publications EP-237362 and EP-329311), allele specific amplification (see, *e.g.*, U.S. Pat. Nos. 5,521,301; 5,639,611; and 5,981,176), mini-sequencing methods, quantitative RT-PCR methods (*e.g.*, the so-called "TaqMan assays"; see, *e.g.*, U.S. Pat Nos. 5,210,015 to Gelfand, 5,538,848 to Livak, *et al.*, and 5,863,736 to Haaland, as well as Heid, C.A., *et al.*, *Genome Research*, 6:986-994 (1996); Gibson, U.E.M, *et al.*, *Genome Research* 6:995-1001 (1996); Holland, P. M., *et al.*, *Proc. Natl. Acad. Sci. USA* 88:7276-7280 (1991); and Livak, K.J., *et al.*, *PCR Methods and Applications* 357-362 (1995)), and various single base pair extension (SBPE) assays.

A number of SPBE assays have been developed, but the general approach is quite similar. Typically, these assays involve hybridizing a primer that is complementary to a target nucleic acid such that the 3' end of the primer is immediately 5' of the variant site or is adjacent thereto. Extension is conducted in the presence of one or more labelled non-extendible nucleotides that are complementary to the nucleotide(s) that occupy the variant site and a polymerase. The non-extendible nucleotide is a nucleotide analog that prevents further extension by the polymerase once incorporated into the primer. If the added non-extendible nucleotide(s) is(are) complementary to the nucleotide at the variant site, then a labelled non-extendible nucleotide is incorporated onto the 3' end of the primer to generate a labelled extension product. Hence, extended primers provide an indication of which nucleotide is present at the variant site of a target nucleic acid. Such methods and related methods are discussed, for example, in U.S. Patent Nos. 5,846,710; 6,004,744; 5,888,819; 5,856,092; and 5,710,028; and in PCT publication WO 92/16657.

SBPE methods involve the addition of labelled nucleotides of high purity which contain low levels of unlabelled nucleotide. Such additions are deemed necessary to obtain acceptable signal-to-noise ratios. The high concentration of label, however, can interfere with the fidelity and rate at which polymerases incorporate non-extendible nucleotide into the primer. Consequently, SBPE methods typically are limited to using polymerases that are less error prone than other polymerases. Typically, the less error prone

enzyme is a non-thermostable enzyme; utilization of this type of enzyme can negatively affect the signal-to-noise ratio, since the non-thermostable polymerases cannot be used in cycling methods to increase the amount of labelled products. In addition, it is often necessary to use a specially modified polymerase to efficiently incorporate a labelled nucleotide analog; failure to use such enzymes can lead to serious levels of misincorporation (see *e.g.*, Reeves and Fuller, *Nature* 376: 796-797 (1995)). Further, because SBPE methods are limited to methods involving high concentrations of labelled nucleotides, the methods are quite expensive.

SUMMARY OF THE INVENTION

The present invention provides methods and kits for detecting and identifying the nucleotide present at a variant site of a target nucleic acid of interest. The methods and kits can be utilized in research, clinical and laboratory settings. In general, the methods involve conducting extension reactions in the presence of a mixture of labelled and unlabelled forms of one or more nucleotides to generate labelled extension products that are characteristic of the nucleotide occupying the variant site of a target nucleic acid. The detection of incorporation of labelled nucleotide provides an indication of the identity of the nucleotide at the variant site since the incorporated nucleotide is complementary to the base at the site of variation. The methods can be used in conducting genotyping analyses and can be performed in multiplexing formats. The methods and kits have utility in diverse applications including, for example, analyzing point mutations and single nucleotide polymorphisms, detection of pathogens, paternity disputes, prenatal testing and forensic investigations.

Certain methods of the invention are methods for analyzing a site of variation in a target nucleic acid and involve contacting a sample of the polynucleotide with multiple copies of a primer whose 3' end hybridizes adjacent to but not including the site of variation in the presence of a mixture of labelled and unlabelled forms of a nucleotide. The unlabelled form is typically 30% to 95% of the total amount of the labelled and unlabelled forms on a molar basis. The primer copies and target polynucleotide are contacted under conditions in which if the nucleotide is complementary to the base occupying the site of variation in the target polynucleotide, at least one copy of the primer is extended by incorporation of the labelled nucleotide to form labelled extension product. The incorporation of the labeled

nucleotide into the primer is then detected; the identity of the base incorporated into the primer indicates which base occupies the site of variation in the target polynucleotide. The nucleotide provided as a mixture can be an extendible nucleotide such as the naturally occurring deoxynucleotides (*e.g.*, dATP, dTTP, dCTP and dGTP), a non-extendible nucleotide (*e.g.*, a dideoxynucleotide or an arabinoside triphosphate) or a combination thereof.

Utilization of a mixture of labelled and unlabelled forms can reduce the incidence of misincorporation that occurs when substantially pure labelled nucleotides are used to conduct extension reactions with thermostable polymerases. Misincorporation with some labelled nucleotides can be sufficiently problematic with certain prior art methods that extension reactions must be performed with non-thermostable polymerases that are less prone to misincorporate labelled nucleotides as compared to thermostable polymerases. The use of non-thermostable polymerases, however, prevents one from conducting thermocycling reactions to increase extension product concentrations since such thermocycling reactions require a thermostable polymerase. By using a mixture of labelled and unlabelled forms, in some instances the level of misincorporation is sufficiently reduced that thermostable and modified thermostable polymerases can be utilized, thus making certain methods of the invention amendable to thermocycling techniques to increase extension product concentration.

Multiple different nucleotides can be included in the extension reaction mixtures in certain methods. For example, in conducting genotyping experiments of a biallelic variant site, the two nucleotides complementary to the bases potentially occupying the variant site of the target nucleic acid can be added, each nucleotide added as a mixture of labelled and unlabelled forms. Thus, certain methods of the invention include contacting a sample containing multiple copies of a target polynucleotide with a plurality of copies of a primer whose 3' end hybridizes adjacent to but not including the site of variation in the presence of a plurality of mixtures of labelled and unlabelled forms of a nucleotide, each mixture containing a different nucleotide. The unlabelled form added is typically 30% to 95% of the combined amount of the labelled and unlabelled forms on a molar basis. The target polynucleotide, primers and multiple mixtures of nucleotides are contacted under conditions in which if the nucleotide of a mixture is complementary to the base occupying the site of variation in the target polynucleotide, at least one copy of the primer is extended by

incorporation of the labelled nucleotide. Incorporation of labelled nucleotide(s) into the primers is then detected, the particular labelled nucleotide(s) incorporated being an indication of the nucleotides present at the variant site.

The invention also provides multiplexing methods in which multiple variant sites can be analyzed at the same time. Certain such methods involve conducting a plurality of template-dependent extension reactions with a plurality of different primers and one or more mixtures of labelled and unlabelled forms of a nucleotide. The different primers utilized hybridize adjacent different sites of variation on target polynucleotides and the different mixtures contain different nucleotides. The unlabelled form in each mixture is typically 30% to 90% of the combined amount of the labelled and unlabelled forms on a molar basis. Each extension reaction involves contacting a sample of the target polynucleotides with multiple copies of one of the different primers, wherein the primer bears an optional label and the 3' end of the primer hybridizes adjacent to but not including the site of variation in the target polynucleotide. The primer copies are exposed to one or more of the mixtures of labelled and unlabelled nucleotides, under conditions such that if the nucleotide of the mixture is complementary to the base occupying the site of variation in the target polynucleotide, at least one copy of the primer is extended by incorporation of the labelled nucleotide.

The various extension reactions generate a plurality of different extension products. Extension products are detected by detecting the incorporation of labelled nucleotides into the extended primers as an indication of the bases occupying the site of variation in the target polynucleotides. Extension products generated from different sites of variation can be distinguished on the basis of the optional label borne by the extended primer and/or label borne by incorporated nucleotide. The extension reactions can be conducted concurrently in a single reaction vessel or in separate reaction vessels. When the different sites of variation are different sites on the same target polynucleotide or different sites on different polynucleotides, the extension reactions can be performed concurrently in a single reaction vessel.

Different extension products can be distinguished by using differentially labelled primers, differentially labelled nucleotides and/or different sized primers for the different variant sites. When the same site of variation on different target nucleic acids is to be examined (e.g., the same SNP site for different individuals), extension reactions for

samples from different individuals are normally performed in separate reaction vessels. The extension products can then be combined and analyzed to detect incorporation of label into the extension products. In this instance also, extension products from different reaction vessels can be distinguished by utilizing differentially labelled primers, differentially labelled nucleotides and/or different sized primers for the different extension reactions.

The invention also provides kits for conducting the methods of the invention. In general, such kits typically include a mixture of labelled and unlabelled nucleotides, the unlabelled nucleotide being 30% to 95% of the combined amount of unlabelled and labelled forms of the nucleotide on a molar basis. The kits can also include a primer that hybridizes to a segment of the target nucleic acid, wherein the 3' end of the primer hybridizes adjacent to the site of variation in the target nucleic acid and bears an optional label, and a polymerase that catalyzes template-dependent extension of a primer in the presence of the appropriate template and nucleotides.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates the major steps of certain extension methods of the present invention.

FIG. 2 depicts a sequence of the amplimer containing the Chr22.53 SNP (A/G polymorphism). The SNP site is represented by the boldfaced "R" in parentheses, indicating that the base at the variant site is A or G. The primer binding sites are shown, with the forward and reverse PCR primer binding sites in bold type and the extension primer binding site underlined.

FIG. 3A shows electropherograms as plots of fluorescence intensity versus time for genotyping experiments conducted for 7 different individuals at the Chr.22.53 SNP site. Experiments were conducted using 200 nM TAMRA-ddATP and 100 nM R110-ddGTP. Signals from extension products bearing the label R110 generated from the G allele are indicated by traces labelled G; signals for extension products bearing the label TAMRA corresponding to the extension product generated from the A allele are shown by traces labelled A. The genotypes, as listed on the left side of each panel, were confirmed by sequencing the amplimers directly.

FIG. 3B provides electropherograms as plots of fluorescence intensity versus time for genotyping experiments at the Chr.22.53 SNP site performed with the same 7

individuals used to generate the results summarized in FIG. 3A. Unlike the results shown in FIG. 3A, however, experiments were conducted with mixtures of labelled and unlabelled nucleotides. The genotypes were determined using a mixture of 15 nM R110-ddGTP/85 nM ddGTP and a mixture of 30 nM TAMRA-ddATP/70 nM ddATP. Extension product bearing the R110 label is indicated by traces labelled G and represents the G allele. Extension product bearing the TAMRA label is represented by traces labelled A and corresponds to the A allele. The correct genotypes are listed on the left side of the panels. Comparison of FIG. 3A and FIG. 3B indicate that equitable genotyping results are obtained in the presence of a mixture of labelled and unlabelled nucleotides.

FIG. 4A depicts electropherograms as plots of fluorescence intensity versus time for genotyping experiments conducted with 7 individuals at the Chr.22.53 site. Experiments were conducted using 100 nM R6G-ddATP and 100 nM R110-ddGTP (*i.e.*, nucleotides were not mixtures of labelled and unlabelled forms). Signals from extension products bearing the R6G label generated from the A allele are represented by traces labelled A. Extension products bearing the R110 label formed from the G allele are represented by traces labelled G. The correct genotypes, as confirmed by sequencing, are listed on the left of the panels. Note the misincorporation of R110-ddGTP into the "AA" homozygotes.

FIG. 4B presents electropherograms as plots of fluorescence intensity versus time for genotyping experiments conducted with the same 7 individuals as described in FIG. 4A at the Chr.22.53 SNP site. Experiments were performed using a mixture of 15 nM R110-ddGTP/85 nM ddGTP and a mixture of 15 nM R6G-ddATP/85 nM ddATP. Extension products labelled with R110 (corresponding to product from the G allele) are indicated by the traces labelled G. Extension products labelled with R6G (corresponding to product from the A allele) are represented by the traces labelled A. Note the suppression of R110-ddG misincorporation into the "AA" homozygotes. The correct genotypes are listed on the left of the traces.

FIG. 5 shows the sequence of the amplimer containing the CYP4.2D6.G4268C SNP (a G/C polymorphic site) (Sachse *et al.*, *Am. J. Hum. Genet.* 60: 284-295 (1997)). The SNP site is indicated by the boldfaced "S" indicating that the variant site is occupied by G or C. The forward and reverse PCR primer binding sites are in bold type and the binding site for the detection primer is underlined.

FIG. 6A presents electropherograms as plots of fluorescence intensity versus time for genotyping experiments conducted with 7 different individuals at the CYP4.2D6.G4268C SNP site. Experiments were conducted using a mixture of labelled and unlabelled nucleotides, specifically 10 nM R110-ddCTP/ 90 nM ddCTP and a mixture of 30 nM TAMRA-ddGTP/ 70 nM ddGTP. Signals from extension products incorporating R110 label and corresponding to the C allele are represented by the traces labelled with a C. Extension products labelled with TAMRA and corresponding to the G allele are represented by traces labelled G. The genotypes, as indicated in parentheses, were also confirmed by sequencing the amplimers directly.

FIG. 6B depicts electropherograms as plots of fluorescence intensity versus time for genotyping experiments conducted with samples from the same 7 individuals described in FIG. 6A at the CYP4.2D6.G4268C SNP site. Experiments were conducted with a mixture of 10 nM R110-dCTP/90 nM CTP and mixtures of 30 nM TAMRA-ddGTP/70 nM ddGTP. Extension products incorporating R110 and generated from the C allele are represented by traces labelled C. Extension products labelled with TAMRA and generated from the G allele are represented by traces labelled G. The correct genotypes are as indicated in parentheses. A comparison of FIG. 6A and 6B demonstrates that a deoxynucleotide can also be used in the primer extension reaction to obtain the correct genotypes.

FIG. 7 depicts a sequence of the amplimer containing the CYP450.2D6.G1749C SNP (G/C polymorphism). The SNP site is represented by the boldfaced "S" indicating that the base at the variant site is G or C. The primer binding sites are shown with the forward and reverse PCR primer binding sites in bold type and the primer binding site underlined.

FIG. 8A shows electropherograms as plots of fluorescence intensity versus time for genotyping experiments conducted for 7 different individuals at the CYP450.2D6.G1749C site. Experiments were conducted using 100 nM R110-ddCTP and 100 nM R6G-ddGTP. Signals from extension products bearing the label R110 generated from the C allele are indicated by traces labelled C. Fluorescent signals for extension products bearing the R6G label generated from the G allele are shown by traces labelled G. The genotypes, as listed on the left side of each panel, were confirmed by sequencing the amplimers directly. As can be seen in this figure, the genotype was correctly identified for each of the 7 individuals.

FIG. 8B provides electropherograms as plots of fluorescent intensity versus time for genotyping experiments at the CYP450.2D6.G1749C site conducted with the same 7 individuals used to generate the results summarized in FIG. 8A. Unlike the results shown in FIG. 8A, however, experiments were conducted with a mixture of labelled and unlabelled nucleotides. Extension reactions were conducted in the presence of a mixture of 10 nM R110-ddCTP/90 nM ddCTP and a mixture of 10 nM TAMRA-ddGTP/90 nM ddGTP. Extension product bearing the R110 label is indicated by traces labelled C and represents extension product formed from the C allele. Extension product bearing the R6G label is represented by traces labelled G and corresponds to the G allele. The correct genotypes are listed on the left side of the panels. A comparison of FIG. 8A and FIG. 8B shows that equitable genotyping results are obtained in the presence of a mixture of labelled and unlabelled nucleotides.

DETAILED DESCRIPTION

I. Definitions

A "nucleic acid" is a deoxyribonucleotide or ribonucleotide polymer in either single or double-stranded form, including known analogs of natural nucleotides unless otherwise indicated.

A "polynucleotide" refers to a single- or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases.

An "oligonucleotide" is a single-stranded nucleic acid typically ranging in length from 2 to about 500 bases. Oligonucleotides are often synthetic but can also be produced from naturally occurring polynucleotides. Oligonucleotides can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphotriester method of Narang *et al.*, *Meth. Enzymol.* 68:90-99 (1979); the phosphodiester method of Brown *et al.*, *Meth. Enzymol.* 68:109-151 (1979); the diethylphosphoramidite method of Beaucage *et al.*, *Tetrahedron Lett.* 22:1859-1862 (1981); and the solid support method described in U.S. Patent No. 4,458,066.

A "primer" is a single-stranded oligonucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions (*i.e.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable

temperature. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template. The term "primer site" or "primer binding site" refers to the segment of the target DNA to which a primer hybridizes. The term "primer pair" means a set of primers including a 5' "upstream primer" that hybridizes with the complement of the 5' end of the DNA sequence to be amplified and a 3' "downstream primer" that hybridizes with the 3' end of the sequence to be amplified.

A primer that is "perfectly complementary" has a sequence fully complementary across the entire length of the primer and has no mismatches. The primer is typically perfectly complementary to a portion (subsequence) of a target sequence. A "mismatch" refers to a site at which the nucleotide in the primer and the nucleotide in the target nucleic acid with which it is aligned are not complementary.

The term "substantially complementary" means that a primer is not perfectly complementary to its target sequence; instead, the primer is only sufficiently complementary to hybridize selectively to its respective strand at the desired primer-binding site.

Hybridizations are usually performed under stringent conditions that allow for specific binding between an oligonucleotide and a target nucleic acid. A stringent condition is any suitable buffer concentration and temperature that allow specific hybridization of the oligonucleotide to a complementary nucleic acid. The phrase "hybridizing specifically to" and related phrases, refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture. See M. Kanehisa, *Nucleic Acids Res.*, 12:203 (1984). Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the oligonucleotides complementary to the target sequence hybridize to the target sequence at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the

temperature is at least about 30 °C for short oligonucleotides (*e.g.*, 10 to 50 nucleotides) and at least about 60 °C for longer probes (*e.g.*, greater than 50 nucleotides). Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide.

A "site of variation" or "variant site" when used with reference to a nucleic acid broadly refers to a site wherein the identity of nucleotide at the site varies between nucleic acids that otherwise have similar sequences. For double-stranded nucleic acids, the variant site includes the variable nucleotide on one strand and the complementary nucleotide on the other strand. A variant site can be the site of a single nucleotide polymorphism or the site of a somatic mutation, including a point mutation, a deletion, an insertion, and a rearrangement, for example.

A "polymorphic marker" or "polymorphic site" is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected population. The first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wild-type form, whereas allelic forms occurring less frequently are referred to as mutant alleles. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic polymorphism has two forms, a tri-allelic polymorphism has three forms and a tetra-allelic polymorphism has four forms.

A "single nucleotide polymorphism" (SNP) occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually preceded by and followed by highly conserved sequences of the allele *(e.g.*, sequences that vary in less than 1/100 or 1/1000 members of the populations). A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele.

The term "naturally occurring" as applied to an object means that the object can be found in nature.

The term "subject" and "individual" are used interchangeably herein to refer to any type of organism, but most typically is used to refer to a human.

II. Overview

The present invention provides a variety of methods, compositions and kits for determining the identity of a nucleotide present at a variant site in a target nucleic acid. The methods are based in part upon the unexpected finding that a mixture of labelled and unlabelled forms for a nucleotide can be used in single base pair extension (SBPE) methods to enhance identification of nucleotides at a site of variation in several respects while reducing the costs of analysis. More specifically, the methods of the invention involve conducting a template-dependent extension of a primer whose 3' end is hybridized adjacent to a variant site in a target nucleic acid. In some methods, a single base complementary to the nucleotide occupying a variant site is integrated into the primer and is a non-extendible nucleotide whose incorporation prevents further extension of the primer. This non-extendible nucleotide is added to the extension reaction as a mixture of labelled and unlabelled forms. In certain methods, other nucleotides, such as an extendible nucleotide or nucleotide analog, can also be added as mixtures of labelled and unlabelled forms. For each nucleotide added as a mixture, typically the unlabelled nucleotide is at least 30%-50% of the total concentration of the nucleotide (*i.e.*, the sum of the labelled and unlabelled forms for that nucleotide) on a molar basis.

SBPE methods that utilize high purity labelled nucleotides suffer in that the fidelity of nucleotide incorporation is negatively affected because labelled nucleotides tend to make many polymerases more error prone (see, *e.g.*, Lee, L.G., *et al.*, *Nucleic Acids Res.* 20:2471-2483 (1992); and Ansorge, W., *et al.*, *Methods Mol. Biol. U.S.A.* 23:317-356 (1993)). An example of this phenomenon is illustrated in the example *infra*. As a consequence of this loss in fidelity, some SBPE methods have used non-thermostable polymerases (*e.g.*, Klenow fragment) that are less affected by labelled nucleotides. The use of such polymerases, however, significantly restricts the utility of SBPE reactions because the use of non-thermostable polymerases precludes the use of thermal cycling reactions to increase signal-to-noise ratios. In view of such restrictions, the use of high concentrations of high purity labelled nucleotides has been deemed necessary to achieve satisfactory signal-to-noise ratios. In addition, certain modified or labelled nucleotides are not readily accepted as

substrates by wild type polymerases. To overcome this limitation, some SBPE methods have made use of a modified thermostable DNA polymerase (see, *e.g.*, Ross *et al.*, *Nature Biotechnology* 16: 1347-1351 (1998)).

The present invention, however, has unexpectedly found that by significantly reducing the concentration of the labelled nucleotides through the use of mixtures of labelled and non-labelled forms of a nucleotide that extension reactions such as SBPE reactions can be performed without sacrificing signal-to-noise ratios. In particular, by using mixtures of labelled and non-labelled forms of a nucleotide, the methods can utilize the more error prone thermostable polymerases without compromising accuracy of nucleotide incorporation, as well as modified thermostable polymerases (*e.g.*, Taq FS). The ability to utilize thermostable polymerases and modified thermostable polymerases specifically selected for incorporating nucleotide analogs means that the methods of the invention are amenable to thermal cycling reactions to increase the amount of extension product formed, thus enabling the methods to achieve high signal-to-noise ratios despite using less labelled nucleotide. Hence, with the methods of the invention, one can obtain good signal-to-noise ratios but with less label and at lower cost. Additionally, the employment of modified thermostable polymerases allows the use of a wider selection of labels since such polymerases can function properly with a wide variety of dye types. Other non-modified polymerases are unable to incorporate many modified nucleotides.

Moreover, the methods of the invention provide greater flexibility in controlling signal intensity which can greatly simplify data analysis, especially for multiplex reactions in which multiple polynucleotide targets and/or variant sites are interrogated simultaneously. Methods employing high concentrations of labelled nucleotides are poorly suited for regulating signal intensities. For example, in multiplex analyses using nucleotides that bear different labels, some signals can be quite weak, whereas other signals can be quite strong. Because the methods of the invention utilize mixtures of labelled and unlabelled forms for a nucleotide, the relative concentration of different labelled nucleotides can be varied to achieve signal output wherein the signal intensities are more consistent. Such consistency facilitates data collection and increases the accuracy of the results of data analysis.

III. Determination of Nucleotide at Variant Site

A. General Description

The methods of the invention generally involve the use of mixtures of labelled and unlabelled forms of at least one nucleotide in conducting primer extension reactions to determine the identity of a nucleotide at a variant site in a target polynucleotide. Although, a variety of primer extension methods have been developed, the general approach of such methods is nonetheless quite similar. In brief, the methods involve hybridizing a primer that is complementary to a target nucleic acid such that the 3' end of the primer hybridizes adjacent to, but does not span, the variant site of the target nucleic acid of interest. The hybridization is typically performed in the presence of one or more labelled nucleotides complementary to a nucleotide that potentially occupies the variant site. Hybridization is performed under conditions allowing primer extension if a nucleotide complementary to a base occupying the variant site in the target nucleic acid is present. Extension results in the incorporation of a labelled nucleotide, thereby generating a labelled extended primer. Extended primers are detected and provide an indication of which base(s) occupy the site of variation in the target polynucleotide. Examples of such primer extension methods include, but are not limited to, U.S. Pat. Nos. 5,710,028 to Eyal *et al.*, 5,856,092 to Dale *et al.*, 5,846,710 to Bajaj *et al.*, 5,888,819 and 6,004,744 both to Goelet *et al.*, and PCT publication WO 92/16657 to Livak, *et al.*, each of which is incorporated by reference in its entirety. The methods of the invention are generally applicable to these methods and other related primer extension methods.

B. Strand Separation and Annealing

1. Strand separation

The various major steps of certain methods of the invention are illustrated in FIG. 1. It will be appreciated by those skilled in the art that the identity of the nucleotides represented at the variant site in FIG. 1 have been arbitrarily selected and that the nucleotides at the variant site could include any of the naturally occurring four bases (*i.e.*, A, T, G or C). The particular variant site shown if FIG. 1 is for an A polymorphic form.

The methods of the invention begin with the treatment of a sample that includes a duplex target nucleic acid to obtain unpaired nucleotides that at least span the variant site of interest or, alternatively, to obtain separate strands. Of course, if the target nucleic acid is already single-stranded, such a step is unnecessary. The term "target nucleic

acid" as used herein refers to single- or double-stranded nucleic acids that include at least one of the variant sites being interrogated. For double-stranded nucleic acids, the variant site includes the nucleotide at the site being examined and the complementary nucleotide in the complementary strand. If a double-stranded target nucleic acid is denatured to form two single strands, each strand can be considered a target nucleic acid and either strand can serve as a template in the methods of the invention.

Strand separation can be achieved using various denaturing conditions that are known in the art including, for example, heat, alkali, formamide, urea, glyoxal and combinations thereof. Typically, strand separation is achieved using heat denaturation at temperatures ranging from 80 °C to about 105 °C for time periods ranging from about 1 to 10 minutes. Alternatively, single-stranded template can be generated through degradation of one strand by exonucleases (see, *e.g.*, Somers et al, Biochimica et Biophysica Acta 1379: 42-52 (1998); Nikiforov et al, PCR Methods and Applications 3: 285-291 (1994); Higuchi and Ochman, Nucleic Acids Research 17: 5865 (1989); and Straus and Zagursky, Biotechniques 10: 376-384 (1991)).

2. Annealing

A primer (also referred to as a detection primer) is then annealed under hybridizing conditions to a template strand of the target nucleic acid. The primer is capable of specifically hybridizing to a segment of the target nucleic acid such that its 3' end is adjacent to the variant site on the target nucleic acid (see FIG. 1). As used herein, the term "adjacent to" when used in reference to hybridization between the primer and target nucleic acid typically means that the primer hybridizes to the target nucleic acid so that its 3' end is immediately 5' to the variant site. However, the 3' end can be located several nucleotides 5' to the variant site so long as none of the nucleotides between the 3' end of the primer and the variant site are the same as the nucleotide that potentially occupies the variant site.

3. Primers

A variety of different types of primers can be utilized with the present methods. Suitable primers include, for example, an oligodeoxyribonucleotide, an oligoribonucleotide, a peptide nucleic acid or a copolymer thereof. Primers can be either naturally occurring nucleic acids or prepared using synthetic methods. If synthesized, the

primers can be synthesized either enzymatically *in vitro*, enzymatically *in vivo* or non-enzymatically *in vitro*.

Depending upon the nature of the target nucleic acid (see section on samples *infra*) various combinations of primer/target nucleic acid duplexes can be formed. For example, in some methods the template is a deoxyribonucleic acid and the primer is an oligodeoxyribonucleotide, an oligoribonucleotide, or a copolymer thereof. In such instances, a DNA polymerase is utilized to generate a DNA product. In certain other methods, the template is a ribonucleic acid and the primer is an oligodeoxyribonucleotide, an oligoribonucleotide, or a copolymer thereof. Reverse transcriptase can be utilized to form a DNA product. In yet other methods, the template is a deoxyribonucleic acid and the primer is an oligoribonucleotide. Added RNA polymerase can produce an RNA product from such a duplex. Finally, if the template is a ribonucleic acid and the primer an oligoribonucleotide, then an RNA replicase can form an RNA product.

Primers are sufficiently long to specifically hybridize to the appropriate target nucleic acid and to form a stable hybridization complex under the extension reaction conditions. Typically, the primers are 15 to 50 nucleotides in length; in other instances, the primers are 20 to 30 nucleotides long. The length of the primers can be adjusted to be longer or somewhat shorter depending upon the particular sequence to which a primer hybridizes (*e.g.*, primers with a high G/C content typically can be shorter than those with a low G/C content). Most typically, the primers are designed to be perfectly complementary over their entire length with the template strand. However, in certain methods the primers are substantially complementary to the target nucleic acid; mismatches in such instances, however, should not adversely affect the stability of the primer/target nucleic acid hybridization complex.

In certain methods, the primer can include one or more moieties that allow for the affinity separation of the extension product from unincorporated reagents and/or target nucleic acid and/or other nucleic acids in the test sample. Such attachment moieties are described further *infra* in the section on detection. Alternatively, the primer can include a modified nucleotide that facilitates selective labelling following the extension reaction. Hence, the term "label" when used in reference to label borne by a primer, or the term "labelled primer," or reference to a primer bearing a label includes primers that actually bear a label, as well as primers that include modified nucleotides for attachment of label

subsequent to extension. Additionally, the primer can include a label to facilitate detection of extended primer. Suitable labels can be selected from those described below in the section on labeled nucleotides.

C. Extension

With continued reference to FIG. 1, once the primer has been allowed to anneal to a target nucleic acid, the duplex of primer and target nucleic acid is contacted with one to four different nucleotides, with at least one nucleotide being added as a mixture of labelled and unlabelled forms. Generally, the nucleotide chosen is selected to be complementary to a nucleotide potentially at the variant site.

The nucleotide(s) included in extension reactions can be any of the naturally occurring deoxynucleotides (*i.e.* dATP, dGTP, dTTP and dCTP) or derivatives, so long as the nucleotide can be incorporated at the 3' end of a primer in a template-dependent fashion. The nucleotide can be an extendible nucleotide, a non-extendible nucleotide or a combination of extendible and non-extendible nucleotides.

An extendible nucleotide refers to nucleotides to which another nucleotide can be attached at the 3' position of the sugar (*e.g.*, the hydroxyl group in the naturally occurring deoxynucleotides dATP, dTTP, dCTP and dGTP) during primer extension. A "non-extendible nucleotide" refers to nucleotide analogs that once incorporated into the primer cannot be extended further, *i.e.*, there is no 3' hydroxyl group or the 3' hydroxyl group has been modified such that another nucleotide cannot be attached at the 3' position. Thus, suitable non-extendible nucleotides include nucleotides in which the 3' hydroxyl group is substituted with a different moiety such that another nucleotide cannot be joined to the non-extendible nucleotide once incorporated into a primer. Such moieties include, but are not limited to, -H, -SH and other substituent groups. Specific examples of non-extendible nucleotides include dideoxynucleotides and arabinoside triphosphates.

In some methods, a deoxynucleotide and a dideoxynucleotide can be used together, such as in determining the nature of a biallelic variant site via primer extension. In the presence of one or more extendible nucleotides, primer extension can go beyond the variant site depending on the configuration of the experiment. This can be advantageous in certain instances as it provides another criterion for allele calling. For example, if one or more nucleotides immediately 3' of the variant site are the same as the nucleotide at the

variant site, extension can proceed 3' of the variant site to include these nucleotides when reactions are conducted with extendible nucleotides complementary to the nucleotide at the variant site. In this case, different allelic forms can be distinguished on the basis of the identity of the label incorporated into the extension product (assuming differentially labelled nucleotides are used for different allelic forms), as well as the size of the extension product.

When extendible nucleotides are included, a variety of techniques can be used to control the extent of the extension reaction. Such techniques include, controlling polymerase concentration, limiting extension reaction times and conducting extension reactions at low temperatures (*e.g.*, approximately 40 °C to 60 °C for Taq FS).

Of the nucleotide or nucleotides provided, at least one is provided as a mixture of labelled and unlabelled forms. If multiple nucleotides are added to a single reaction, at least one nucleotide is provided as a mixture of labelled and unlabelled forms. The concentration of the unlabelled form for a particular nucleotide relative to the total concentration of labelled and unlabelled forms for that nucleotide is at least 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55% or 60% on a molar basis. The upper limit on the concentration of the unlabelled forms relative to the total amount of both forms is typically 80%, 85%, 90% or 95% on a molar basis. In certain methods, the unlabelled concentration typically is 30% to 95% on a molar basis. In other methods, the unlabelled form is generally 50% to 90% on a molar basis. In still other methods, ratio of unlabelled to labelled forms on a molar basis is 1:2. Although commercially available labelled nucleotides (*e.g.*, TAMRA-ddATP), can sometimes contain relatively low levels of unlabelled forms (typically about 5%), the concentration of unlabelled forms used in the present invention are above such levels and is in accordance with the ranges set forth above.

The nucleotides can be provided according to a variety of different formats. For example, one to four nucleotides can be provided as mixtures of labelled and unlabelled forms within a single reaction mixture. For example, a sample can be divided into different reactions and a different mixture of labelled and unlabelled nucleotides added to each reaction. Thus, for example, if a variant site includes two allelic forms, a sample can be divided into two different reactions. A mixture containing labelled and unlabelled nucleotide complementary to one allelic form can then be added to one reaction mixture and a mixture containing labelled and unlabelled nucleotide complementary to the other allelic form added to the other reaction mixture (*e.g.*, adenosine nucleotide added to one reaction and guanine

nucleotide added to the other for an A/G polymorphism). Alternatively, both mixtures can be added to a single reaction if the nucleotides are differentially labelled.

Adding a single nucleotide to a reaction mixture ensures that only a single type of extension occurs and can in some instances reduce costs. The addition of all 4 nucleotides can in some instances increase the fidelity of incorporation. Also, the addition of all 4 nucleotides can allow samples to be more rapidly analyzed by reducing the number of reactions.

Further, if multiple nucleotides are added to a single reaction, it is not necessary for every nucleotide to include both labelled and unlabelled forms. Thus, for example, if all four nucleotides are added to a reaction mixture (*e.g.*, ddATP, ddTTP, ddCTP and ddGTP), one, two, three or all four of the nucleotides can be provided as mixtures of labelled and unlabelled forms.

As indicated above, in some instances the 3' end of the primer is several nucleotides upstream of the variant site. The nucleotides of the target nucleic acid in this gap should not include a nucleotide that potentially occupies the variant site being analyzed. In such a situation, unlabelled deoxynucleotide triphosphates (dNTPs) complementary to the nucleotides between the 3' end of the primer and the variant site should be included in the extension reactions. These dNTPs are in addition to at least one nucleotide added to the reaction mixture that is complementary to a base potentially at the variant site and added as a mixture of labelled and unlabelled forms.

A polymerase is also added to the reaction mixture to initiate the incorporation of one or more of the nucleotides. If an added nucleotide is complementary to the nucleotide at the variant site of the target nucleic acid, then the polymerase catalyzes the incorporation of this nucleotide to generate an extended primer. Some of the extended primer product is labelled as a consequence of the incorporation of the labelled nucleotide, whereas other extension products lack label because of incorporation of unlabelled nucleotide (see FIG. 1). Since the incorporated nucleotide is complementary to the nucleotide occupying the variant site, the incorporated nucleotide provides the basis for identifying the nucleotide(s) at the site of variation.

As noted above, by providing the nucleotides as mixtures of labelled and unlabelled forms, the present methods can allow for increased fidelity of nucleotide incorporation as compared to methods using primarily labelled nucleotides in which the label

can interfere with incorporation. The enhanced fidelity of incorporation means that the present methods can utilize a wider range of polymerases, including thermostable and modified thermostable polymerases. Methods using primarily labelled non-extendible nucleotides, in contrast, often are limited to less error prone polymerases that are often non-thermostable.

The terms "thermostable polymerase" and "modified thermostable polymerase" refer to those polymerases whose activity is maintained even when reaction mixtures are heated sufficiently to denature duplex nucleic acids to form single-stranded nucleic acids for attachment of primers. Examples of suitable thermostable polymerases include, but are not limited to, Taq FS (Applied Biosystems), ThermoSequenase (Amersham/Pharmacia Biochemicals), and Taq (Applied Biosystems). The modified thermostable polymerases refer to polymerases that have been modified to have increased binding affinity for modified nucleotides (*e.g.*, dideoxynucleotides) and thus readily incorporate such nucleotides. The term "non-thermostable polymerase," in contrast, refers to polymerases which undergo a significant or total loss of activity when heated to achieve denaturation of double-stranded nucleic acids into single-stranded nucleic acids such that the primer can hybridize to its primer binding site. Non-thermostable polymerases include, for example, DNA polymerase I from *E. coli* (*i.e.*, Klenow fragment), Sequenase (USB), and large fragment of Bst Polymerase (New England BioLabs).

The ability to use thermostable polymerases, especially those developed for fluorescent sequencing (*e.g.*, Taq FS and ThermoSequenase), means that the present methods can optionally utilize thermocycling methods to increase the amount of extended primer and allows the use of a greater selection of labeled nucleotides. Because of such capabilities, the present methods can utilize less labelled material without sacrificing signal-to-noise ratios, thereby providing strong signals at lower cost. As shown in FIG. 1A, the thermocycling methods involve repeating the strand separation, annealing and extension reactions with a thermocycler to regulate the temperatures at the appropriate level to promote these steps (see, *e.g.*, Proudfoot, *et al.*, *Science* 209:1329-1336 (1980) and Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory Press (1989)).

D. Detection

Extension product formation is generally detected by detecting the incorporation of label into the primer by various direct or indirect methods. In some instances, extension product is initially separated from unreacted reactants in the extension reaction mixture before extension product is detected. With certain techniques, however, such separation is unnecessary and the methods can be performed in a homogenous format.

When more than one nucleotide is added to extension reactions and each nucleotide is added as a mixture of labelled and unlabelled forms, one can control the signal intensities for extension products incorporating the different labels by adjusting the relative concentration of the different labelled nucleotides. This allows one to regulate signal intensities such that signals for differentially labeled extension products are similar. The term "similar" when used to refer to relative signal strengths means that the difference in signal intensity for a given labelled extension product relative to the average signal intensity for all the extension products formed from the same variant site under similar reaction and detection conditions is typically less than 10%, more frequently less than 30% and most often less than 50%.

1. Separation Based Methods

Once extension reactions have been completed, extended primer product is typically separated from unextended primers and unincorporated labelled nucleotide to facilitate analysis. Following removal of components that might interfere with the detection step, extended primer is analyzed for presence or absence of label. The labelled nucleotide incorporated into a primer is an indicator of the nucleotide that occupies the variant site.

Separation of the extended primers from other reaction components can be achieved in a variety of ways. In one approach, the primer includes an attachment moiety that is one component of an affinity pair and that allows for affinity purification of extended primer from other components of the extension reaction. Typically, the attachment moiety is located at or near the 5' end of the primer. Alternatively, the attachment moiety is connected to the nucleotide. The other member of the affinity pair is frequently attached to a solid support such that extended primer bearing label can be bound to the support via the attached member of the affinity pair. Other reaction components can then be washed away. Another option besides the use of affinity purification is to separate extended primer from other

reaction components using gel electrophoresis. Yet another option is to selectively inactivate the label associated with unincorporated nucleotide.

A variety of different attachment moieties can be used as part of an affinity pair to achieve purification of the extended primer from other components. In general terms, the attachment moiety and the other component of the affinity pair include two agents that are capable of specifically binding to one another. Examples of such binding pairs include, but are not limited to, polynucleotide/complementary polynucleotide, biotin/avidin, antigen/antibody and heavy metal/thiol group. In some instances, one member of the affinity pair is attached to a solid support. A solution containing (or potentially containing) a primer bearing the complementary member of the affinity pair is then contacted with the support. After allowing the two components an opportunity to bind and form a complex, other species in the extension reaction mixture can be washed from the support.

Thus, for example, in one suitable arrangement, the attachment moiety is a polynucleotide that serves as a 5' extension to the primer. A complementary nucleotide is attached to a solid support and is capable of selectively binding the extension primer. Alternatively, an antigen functions as the attachment moiety and an antibody specific thereto is attached to the support. In yet another arrangement, a thiol group is linked to the primer and serves as the attachment moiety. A heavy metal group attached to the solid support can be used to selectively bind the thiolated primer.

The attachment moiety can be attached at any point of the primer where it does not interfere with the extension reaction. Most typically, the attachment moiety is attached at or near the 5' end of the primer. However, in some instances, the attachment moiety is connected to a more internal nucleotide.

Instead of attaching the attachment moiety to the primer, in some methods the attachment moiety is part of the nucleotide. The attachment moiety can be selected from the group of affinity pairs described above, for example (see, also, U.S. Pat. No. 5,710,028 to Eyal, *et al.*). Alternatively, one can obtain antibodies specific to a fluorescent dye label on the nucleotide (e.g., an antibody elicited to fluorescein as a hapten). Such antibodies have been discussed (see, Voss, E.W., Jr. (Ed.) *Fluorescein Hapten: An Immunological Probe*).

A variety of different types of supports can be utilized in methods employing affinity-binding pairs. Suitable supports include, but are not limited to, beads, microparticles, the surface of a microtiter well, a filter and a glass slide. Similarly, the supports can be

formed from any material stable to the binding and washing conditions including, for example, glass, polystyrene, cellulose, latex, nitrocellulose, nylon, polyacrylamide, dextran and agarose.

As an alternative to the use of affinity binding pairs, extended primers can be separated from other reaction components by a variety of size based separation techniques such as gel electrophoresis and size exclusion chromatography (*e.g.*, HPLC). In some methods, separation of components by gel electrophoresis and the detection step (typically detection of fluorescence from a fluorescent label attached to the extended primer via the incorporated nucleotide) is performed using a single integrated instrument, such as the Prizm DNA Sequencers from Applied Biosystems, and MegaBACE from Molecular Dynamics.

2. Homogeneous Assays

Certain methods are conducted in a homogenous assay format in which extension products do not need to be separated from other extension reaction components (*e.g.*, unextended primer and unincorporated nucleotide). In some methods, this is accomplished using donor and acceptor fluorophores, including fluorescence resonance energy transfer pairs. The fluorophores are chosen so that the emission spectrum of one fluorophore (*i.e.*, the donor fluorophore) overlaps the excitation spectrum of the other fluorophore (*i.e.*, the acceptor fluorophore).

With such methods, the labelled nucleotide bears one member of the donor/acceptor dye pair and the other member is attached to the primer. The primer can be labelled at a position such that upon the incorporation of the labelled nucleotide the donor and acceptor are brought into an energy transfer relationship, wherein fluorescence energy can be transferred from the donor to the acceptor. By measuring fluorescence changes that occur as a consequence of energy transfer (*e.g.*, a decrease in the fluorescence intensity of the donor or an increase in the fluorescence intensity of the acceptor), one can detect the incorporation of label onto the primer without having to separate extension product from unreacted reactants. Specific labels suitable for use in such methods are discussed infra in the section on labels. Further guidance on such methods is described in U.S. Patent No. 5,945,283.

3. Indirect Methods

An alternative detection method is to measure the change in the substrate (*i.e.*, free nucleotides) concentrations as an indication of which nucleotide(s) was incorporated into the primer during the extension reaction. Fluorescence polarization techniques can be utilized conveniently for this purpose. This technique is able to distinguish between large and small molecules based on molecular tumbling. Large molecules (*e.g.*, labelled extension product) tumble in solution much more slowly than small molecules. Thus, the signal from a labelled nucleotide incorporated into a primer can be distinguished from the labelled nucleotide free in solution (see, *e.g.*, Chen *et al.*, Genome Research 9:492-8 (1999), and U.S. Pat. No. 5,593,867 to Walker *et al.*, both of which are incorporated by reference in their entirety).

4. Labels

The labelled nucleotide includes a label that is either directly or indirectly detectable. The label covalently attached to the nucleotide, and optionally to the primer, can be any compound or molecule that can be detected and that does not significantly interfere with the extension reaction (*e.g.*, interfering sufficiently such that an undetectable amount of extension product is formed and/or causing elevated rates of misincorporation such that an accurate determination of the identity of the nucleotide at the variant site is not possible). Suitable labels include, but are not limited to, fluorophores, chromophores, molecules that emit chemiluminescence, magnetic particles, radioisotopes, mass labels, electron dense particles, electrochemically active molecules, enzymes, cofactors, substrates for enzymes and ligands having specific binding partners (*e.g.*, avidin/biotin). Mass labels can be prepared from various monomers. By joining differing numbers of monomers together, mass labels of differing molecular weight can be prepared. Essentially, any type of monomer that can be joined together and then attached to a primer can be used, so long as the mass label does not interfere with hybridization of the primer to the target nucleic acid.

Certain methods utilize fluorescent molecules as the labels, as a number of commercial instruments have been developed for the detection of fluorescently labelled nucleic acids. A variety of fluorescent molecules can be used as labels including, for example, fluorescein and fluorescein derivatives, rhodamine and rhodamine derivatives, naphthylamine and naphthylamine derivatives, cyanine and cyanine derivatives, benzamidizoles, ethidiums, propidiums, anthracyclines, mithramycins, acridines,

actinomycins, merocyanines, coumarins, pyrenes, chrysenes, stilbenes, anthracenes, naphthalenes, salicyclic acids, benz-2-oxa-1-diazoles (also called benzofurazans), fluorescamines and bodipy dyes.

For those methods in which the detection primer and/or the detection product are labelled with fluorescent dyes capable of energy transfer to enhance emission intensities or simplify the assay, a number of donor (or reporter) and an acceptor (or quencher) dyes are available. One group of donor and acceptor dyes includes the xanthene dyes, such as fluorescein dyes, and rhodamine dyes. A variety of derivatives of these dyes are commercially available. Often functional groups are introduced into the phenyl group of these dyes to serve as a linkage site to an oligonucleotide. Another general group of dyes includes the naphthylamines which have an amino group in the alpha or beta position. Dyes of this general type include 1-dimethylaminonaphthyl-5-sulfonate, 1-anilino-8-naphthalene sulfonate and 2-p-toluidinyl-6-naphthalene sulfonate.

Other dyes include 3-phenyl-7-isocyanatocoumarin, acridines, such as 9-isothiocyanatoacridine and acridine orange, pyrenes, benzoxadiazoles and stilbenes. Additional dyes include 3-(ε-carboxypentyl)-3'-ethyl-5,5'-dimethyloxa-carbocyanine (CYA); 6-carboxy fluorescein (FAM); 5&6-carboxyrhodamine-110 (R110); 6-carboxyrhodamine-6G (R6G); N',N',N',N'-tetramethyl -6-carboxyrhodamine (TAMRA); 6-carboxy-X-rhodamine (ROX); 2', 4', 5', 7', - tetrachloro – 4 – 7 – dichlorofluorescein (TET); 2', 7' – dimethoxy – 4', 5' – 6 carboxyrhodamine (JOE); 6-carboxy-2'4,4',5',7,7'-hexachlorofluorescein (HEX); ALEXA; Cy3 and Cy5. These dyes are commercially available from various suppliers such as Applied Biosystems Division of Perkin Elmer Corporation (Foster City, CA), Amersham Pharmacia Biotech (Piscataway, NJ), and Molecular Probes, Inc. (Eugene, OR).

Further guidance regarding the selection of donor and acceptor pairs that can effectively be used with the methods of the present invention includes: *Fluorescence Spectroscopy* (Pesce *et al.*, Eds.) Marcel Dekker, New York, (1971); White *et al.*, *Fluorescence Analysis: A Practical Approach*, Marcel Dekker, New York, (1970); Berlman, *Handbook of Fluorescence Spectra of Aromatic Molecules*, 2nd ed., Academic Press, New York, (1971); Griffiths, *Colour and Constitution of Organic Molecules*, Academic Press, New York, (1976); *Indicators* (Bishop, Ed.). Pergamon Press, Oxford, 19723; and Haugland, *Handbook of Fluorescent Probes and Research Chemicals*, Molecular Probes, Eugene (1992).

IV. Multiplexing/Pooling

The fact that the methods of the invention generate labelled products allows for the facile extension of the basic method described above to multiplexing formats in which the identity of a nucleotide at multiple variant sites is determined in a single reaction. Such formats allow for rapid sequence determinations in many loci and/or individuals simultaneously. The multiple sites can be multiple sites on the same target nucleic acid, such sites being within the same gene or at sites in different genes. Alternatively, the multiple sites can be the same site on target nucleic acids obtained from different individuals, or multiple different sites on target nucleic acids from different individuals.

The multiplexing methods closely parallel the methods just described. Primers for each of the different variant sites are annealed to their respective binding sites. The general structure of the primers and the methods for conducting the extension reactions is as set forth above. Here too, the particular labelled nucleotide(s) incorporated into the extended primer serves to identify the nucleotide(s) present at the variant site.

In order to correlate the multiple extension products with the various sites, a number of different strategies can be utilized to aid in determining which extension product corresponds with which variant site. One option is to differentially label the nucleotides such that different labels are used for different variant sites.

Another approach involves using primers of different lengths so that the extension products formed are of differing sizes and capable of being distinguished by size fractionation (*e.g.*, by gel electrophoresis). If this approach is utilized, the primers can differ by as little as a single nucleotide, although typically the difference in size is larger to facilitate the ease by which different extension primers can be distinguished.

Alternatively, primers can be tagged with distinctive identifier tags. A number of different tags can be utilized. In some instances, a secondary label is used in conjunction with the label used to identify the nucleotide present at a particular variant site. Such secondary labels can be any type of molecule or compound that is detectable (see *supra*). In other instances, the tag is part of an affinity pair and facilitates separation of the different extension products one from the other, in addition to functioning to separate the extended primer from other extension reaction components as described above. Suitable affinity pairs include those described *supra* in the section on detection methods.

The methods utilizing primers of different size or tagged primers can be used in conjunction with the scheme in which different labels are used for different variant sites or different nucleotides at a particular variant site. In this way, extension products can be identified and distinguished from one another on the basis of two criteria rather than simply one criterion. For example, in some methods, different nucleotides are attached to different labels. Further, primers include different tags for different variant sites. The different extension products can then be identified and/or separated both according to the different tags and the different labels.

The methods can also be utilized in pooling studies to determine the allele frequency of a variant site in a study population. Typically, in these type of experiments, the DNA samples from different individuals are pooled together. Then the method of this invention can be used to analyze the presence of each allele in the mixed templates. By comparing the signal intensities of each allele with a reference set (for example, the heterozygotes, the homozygotes or a mixture of both at a known ratio), the prevalence of the alleles in the population can be determined. (For a general discussion of pooling studies see, *e.g.*, Breen G. *et al., BioTechniques* 28:464-468 (2000); Risch N. and Teng, J., *Genome Res.* 8:1273-1288 (1998); Shaw, S.H. *et al., Genomer Res.* 8:111-123 (1998); and Scott, D.A. *et al., Am. J. Hum. Genet.* 59:385-391 (1996), each of which is incorporated by reference in its entirety).

V. Genotyping

A diploid organism contains two copies of each gene. Genotyping involves the determination of whether a diploid organism contains two copies of the reference allele (a reference-type homozygote), one copy each of the reference and variant allele (*i.e.*, a heterozygote), or contains two copies of the variant allele (*i.e.*, a variant-type homozygote). When conducting a genotyping analysis, the methods of the invention can be utilized to interrogate a single variant site. However, as described above in the section on multiplexing/pooling, the methods can also be used to determine allelic frequency in a group of individuals, as well as the genotype of an individual in many different DNA loci, either on the same gene, different genes or combinations thereof.

Most typically, SNPs consist of two allelic forms, *i.e.*, the variant site includes one of two different nucleotides. The sample can contain nucleic acids representative of the

two copies of the target nucleic acid of interest. Analyses can be conducted with a single nucleotide, but more typically labelled nucleotides complementary to both nucleotides potentially at the site of variation are utilized. Formation of a single labelled extended primer indicates that the sample is from a homozygote. The particular labelled nucleotide incorporated signifies whether the sample is from a reference-type or variant-type homozygote. The formation of two labelled extension products indicates that the sample is from a heterozygote. Reactions can be conducted separately such that each nucleotide is added to a different reaction mix, or reactions can be conducted in a single reaction mixture containing both nucleotides. If reactions are conducted in a single reaction vessel, the labelled nucleotides are differentially labelled so that the different allelic forms can be distinguished. If different reactions are conducted with each labelled nucleotide, the labels for each labelled nucleotide can be the same or different since the particular nucleotide added to each reaction is tracked.

For polymorphisms that include more than two allelic forms, additional labelled nucleotides can be used. For example, for triallelic polymorphisms, three differentially labelled nucleotides can be used. In like manner, with tetra-allelic polymorphisms, four differentially labelled nucleotides can be employed. Here, too, all the nucleotides can be added to a single reaction mixture or to separate reaction mixtures. Typically, any additional nucleotides are provided as mixtures of labelled and unlabelled forms.

The use of mixtures of labelled and unlabelled forms for any particular nucleotide can significantly enhance data acquisition and analysis during genotyping experiments. A problem with methods that use different labels for different nucleotides is that some labels are considerably weaker or stronger signal emitters than other labels in a particular detection devise. This can be problematic in that if two labels of varying signal strength are used for genotyping a biallelic SNP, heterozygotes can generate two peaks centered at different wavelengths (*i.e.*, two different colors), each having differing heights. This can frustrate an accurate identification of the allelic forms present because misincorporation, even at a low level, can generate a similar pattern. With the methods of the present invention, however, similar signal intensities for labels having different signal strengths, can be easily achieved simply by altering the ratio of labelled to unlabelled forms for each nucleotide.

The ability to use the methods of the invention to make rapid genotyping determinations provides a powerful tool in genetic analysis and ascertaining the susceptibility of an individual to a disease. Individuals that are mutant homozygotes for an allele associated with a particular disease are at higher risk of having the disease than a heterozygote or a homozygote for the other allele. The heterozygote, however, is a carrier of the allele associated with the disease. Such knowledge can be useful in prenatal and other types of medical and genetic counseling, for example.

VI. Samples

A. Types of Target Nucleic Acids

The methods of the present invention can be utilized to determine the identity of a nucleotide at a variety of different types of variant sites including, but not limited to, SNPs and mutations such as transitions, transversions, insertions and deletions. The presence or absence of a target nucleic acid in a sample can be detected generally as the presence or absence of a particular nucleotide at a particular site. Individual nucleotides located at a particular site can also be identified by the methods described herein.

The methods presented are generally applicable to deoxyribonucleic acids, ribonucleic acids, or copolymers thereof. The nucleic acids can be single-stranded or double-stranded. The target nucleic acid can include non-naturally occurring nucleotide analogs including, for example, deoxyinosine or 7-deasa-2-deoxyguanosine. Such analogs destabilize duplex DNA and allow a primer annealing and extension reaction to occur in double-stranded nucleic acids without completely separating the two strands. In some instances, RNA samples are first reversed transcribed to form cDNA before use.

The target nucleic acid can be only a fraction of a larger nucleic acid or can be present initially as a purified and discrete molecule. Additionally, the target nucleic acid can constitute the entire nucleic acid or can be a fraction of a complex mixture of nucleic acids. The target nucleic acid can be synthesized enzymatically *in vivo*, synthesized enzymatically *in vitro*, or synthesized non-enzymatically.

B. Sources

The target nucleic acid can be from any source. The samples that include the target nucleic acids can be natural or synthetic using enzymatic or organic synthesis

techniques. Likewise, the sample can be taken from any organism, including but not limited to, plants, microorganisms (*e.g.*, bacteria, fungi and viruses), vertebrates, invertebrates and mammals (*e.g.*, humans, primates, horses, dogs, cows, pigs and sheep).

For assay of genomic DNA, virtually any biological sample (other than pure red blood cells) is suitable. Samples can be obtained from the tissues or fluids of an organism or organisms; samples can also be obtained from cell cultures, tissue homogenates or synthesized as described above. For example, samples can be obtained from whole blood, serum, semen, saliva, tears, urine, fecal material, sweat, buccal, skin, spinal fluid and hair. Samples can also be derived from *in vitro* cell cultures, including the growth medium, recombinant cells and cell components. For assay of cDNA or mRNA reverse transcribed to form cDNA, the tissue sample is obtained from an organ in which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source. Samples for use in prenatal testing can be obtained from amniotic fluid.

The target nucleic acid(s) can also be obtained from non-living sources suspected of containing matter from living organisms. For example, in the instance of samples obtained for forensic analysis, the target nucleic acids can be obtained from samples of clothing, furniture, weapons and other items found at a crime scene.

C. Sample Preparation

In some instances, the samples contain such a low level of target nucleic acids that it is useful to conduct a pre-amplification reaction to increase the concentration of the target nucleic acids. If samples are to be amplified, amplification is typically conducted using the polymerase chain reaction (PCR) according to known procedures. See generally, *PCR Technology: Principles and Applications for DNA Amplification* (H.A. Erlich, Ed.) Freeman Press, NY, NY (1992); *PCR Protocols: A Guide to Methods and Applications* (Innis, *et al.*, Eds.) Academic Press, San Diego, CA (1990); Mattila *et al.*, *Nucleic Acids Res.* 19: 4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1: 17 (1991); *PCR* (McPherson *et al.* Ed.), IRL Press, Oxford; and U.S. Patent Nos. 4,683,202 and 4,683,195, each of which is incorporated by reference in its entirety. Other suitable amplification methods include: (a) ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4:560 (1989) and Landegren *et al.*, *Science* 241:1077 (1988)); (b) transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86:1173 (1989)); (c) self-sustained sequence replication

(Guatelli *et al., Proc. Natl. Acad. Sci. USA*, 87:1874 (1990)); and (d) nucleic acid based sequence amplification (NABSA) (see, Sooknanan, R. and Malek, L., *Bio Technology* 13: 563-65 (1995)), each of which is incorporated by reference in its entirety.

Further guidance regarding nucleic sample preparation is described in Sambrook, *et al., Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, (1989), which is incorporated herein by reference in its entirety.

VII. Kits/Compositions

Compositions and kits for conducting the sequence and genotyping determinations described herein are also provided by the invention. The compositions include mixtures of labelled and unlabelled nucleotides such as those described above. The concentration of labelled to unlabelled forms for a nucleotide is as indicated above, but most typically the unlabelled form is 50% to 90% of the total concentration of the labelled and unlabelled forms as expressed on a molar basis. Although various nucleotides can be utilized in the kits, most frequently the nucleotide is a dideoxynucleotide. The labelled forms can include any of the labels described *supra*; most typically, however, the label is a fluorophore, especially FAM, ROX, TAMRA, R110, R6G, JOE, TET, HEX, Alexa dyes, Cy3 and Cy 5.

The compositions of the invention can be combined with other components to form kits. Typically, the kits include one or more of the primers. Once hybridized to a target nucleic acid of interest, the 3' end of the primer is adjacent the variant site of the target nucleic acid of interest. Often the primers are designed for use in detecting one or more SNPs such as those described *infra*. The number of primers included in the kit can vary. Generally, the kits include at least 2, 3, 4 or 5 primers. Other kits can include more primers, such as at least 10, 15, 20 or 25 primers. In some kits, the primers can also bear a label or attachment moiety to facilitate the detection/separation process.

The kits can include various other components for conducting template-dependent extension reactions including, for example, a polymerase, nucleoside triphosphates, and buffers. The polymerase can be either a thermostable or non-thermostable polymerase. Kits can also include the necessary electrophoretic components to size separate the extension products formed during an analysis. Such components include gel polymers (e.g., agarose), polymerizing agents and buffers. Typically, the kits also include containers

for housing the various components and instructions for using the kit components to conduct an analysis.

VIII. Utility

The methods, compositions and kits of the invention are generally useful for determining the identity of a nucleotide at a variant site. These methods, however, find use in a variety of more specific applications. One use is the identification and detection of point mutations (*e.g.*, somatic point mutations), specifically those mutations correlated with diseases. For example, the methods described herein are useful for identifying whether a nucleic acid from a particular subject includes a wild-type allele or a mutant allele at a particular SNP site. Furthermore, the methods can be utilized to establish the genotype of the individual being tested (*i.e.*, distinguish whether the individual is a reference-type homozygote, a heterozygote or a variant-type homozygote).

The genotyping utility of the present methods makes them useful within the context of medical diagnosis and prognosis. Since many SNPs are associated with various diseases, clinicians can utilize the results of the genotype study to assess the presence of disease, whether an individual is a carrier of disease, the likelihood that an individual will get a particular disease and the likely efficacy of various treatment alternatives.

The methods also have a variety of non-medical uses. Such utilities include detecting pathogenic microorganisms, paternity testing and forensic analysis. The methods can also be used to identify SNPs in non-humans, including, for example, other animals, plants, bacteria and viruses.

These various uses are described more fully below.

A. Correlation Studies

Use of the methods of the present invention to acquire diagnostic information involves obtaining a sample from a number of different individuals known to have a common disease and conducting screening tests to determine whether they consistently share a common genotype at one or more SNP sites. The results of such screening can be used to establish correlations between certain genotypes and certain diseases.

In a related fashion, the methods of the invention can be used to develop correlations between certain genotypes and patient prognosis. For example, the genotype of a

population of individuals suffering from a common disease can be determined at one or more SNP sites. The health history of the individuals can be monitored with time to establish correlations between certain genotypes and disease outcomes.

The methods of the invention can also be used to formulate optimal treatment protocols for a particular disease. For example, the ability of an individual to metabolize certain drugs may be associated with a particular genotype(s). The methods described herein can be used to place individuals into groups that share a common phenotype and genotype. The group can then be subdivided into various groups that each receives different forms of treatment. By monitoring the health status of the different treatment groups over time, the most effective treatment program for a particular genotype can be established.

B. Use of Current Methods as Screening and Therapeutic Tool

In instances in which a correlation between a particular genotype and disease state or drug response have already been established, the methods of the invention can be utilized as a diagnostic tool, a prognostic tool and as a means for optimizing treatment.

For patients having symptoms of a disease, the methods of the present invention can be used to determine if the patient has a genotype known to be associated with a disease that commonly causes the symptoms the patient exhibits. For example, if the genotyping methods of the invention show that the individual has a genotype associated with a particular disease and further that the genotype is associated with poor recovery (*e.g.*, a variant-type homozygote), the physician can counsel the patient regarding the likely effectiveness of aggressive treatment options and the option of simply foregoing such treatments, especially if the disease is quite advanced. On the other hand, if the genotype is associated with good recovery, the physician can describe a range of treatment options varying from simply monitoring the disease to see if the condition worsens or more aggressive measures to ensure that the disease is attacked before it gets worse.

The methods of the present invention are also valuable for assessing the actual risk of an individual known to be susceptible to acquiring a disease (*e.g.*, an individual coming from a family that has a history of suffering from a disease). By determining whether the individual is a homozygote for the SNP associated with the disease or a heterozygote, a physician can more accurately assess and counsel the patient regarding the likelihood that the

patient will begin suffering from disease, factors involved in triggering the disease and the pros and cons regarding different treatment alternatives.

Similarly, certain methods of the invention can also be used to identify individuals at risk for disease, even though they have no symptoms of disease or no known susceptibilities to disease. An individual in this category would generally have no disease symptoms and have no family history of disease. In such cases, the methods of the present invention can be utilized as a useful preventive screening tool. Using the methods of the present invention, a number of selected SNP sites known to be associated with certain diseases can be interrogated to identify the genotype of the individual at those sites. If a particular genotype were identified that was known to be associated with a particular disease, then a physician could advise the individual regarding the likelihood that the disease would manifest itself and the range of treatment options available.

In yet another application, the present invention can be useful in formulating optimal treatment for a patient. For example, an individual's response to a drug may be predicted based on his or her genotype. The information enables a physician to prescribe the most effective medication with minimal side effects.

C. Examples of Diseases that can be Monitored

A large number of diseases have been shown to be correlated with particular allelic forms of SNPs. A large number of such SNPs are listed in WO 93/02216 and by Cooper *et al.* (*Hum. Genet.* 85:55-74 (1990)), both of which are incorporated herein by reference in their entirety. Specific examples of diseases associated with SNPs include: sickle cell anemia and β-thalassemias (mutation in β-globin gene; Antonarakis, *New Eng. J. Med.*, 320:153-163 (1989)), cystic fibrosis (mutation in cystic fibrosis transmembrane receptor (CFTR); see Kerem, *et al.*, *Science* 245:1073-1080 (1989)), hyperlipoproteinemia (mutation in apolipoprotein E gene; see Mahley, *Science* 240:622-630 (1988)), a wide variety of autoimmune diseases (mutations in human major histocompatibility complex; see Thomson, *Ann. Rev. Genet.*, 22:31-50 (1988); Morel *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:8111-8115 (1988); and Scharf, *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:3504-3508 (1988)) and the formation of oncogenes (mutations to the human ras-gene family; see, *e.g.*, Bos *et al.*, *Nature*, 315:726-730 (1985); Farr *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:1629-1633 (1988); and Neri, *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:9268-9272 (1988)). Other genes containing

SNPs associated with disease include genes encoding for angiotensinogen, angiotensin converting enzyme, cholesterol ester transfer protein, dopamine receptors, serotonin receptors, and HIV reverse transcriptase (RT).

D. Other Uses

The methods described herein can also be used to identify point mutations in pathogens that could potentially result in altered pathogenicity or resistance to certain therapeutics. The methods can also be used to identify cells and strains having a desired genetic constitution for use in various biotechnology applications. The methods described herein can also detect the presence of somatic mutations that can result in various diseases, including cancer for example.

With knowledge gained from the genotyping methods described herein, clinicians can conduct prenatal testing using cells obtained from a fetus to check for a variety of inheritable diseases, such as those diseases associated with the SNPs listed above. The methods can also be used to identify carriers of mutant alleles. Such information can be of use by a couple prior to conception as they evaluate the risks of having a child with certain birth defects or inheritable diseases.

Methods of the invention can also be utilized in various identification applications, such as in the field of forensic medicine or paternal testing. In the case of forensic analysis, polymorphisms in specific genes can be determined in, for example, blood or semen obtained from a crime scene to indicate whether a particular suspect was involved in the crime. In like manner, polymorphism analysis may be utilized in disputes to aid in determining whether a particular individual is the parent of a certain child.

In another application, certain methods of the invention are used in blood typing or tissue classification. Tissue classifications, for example, can be determined by identifying polymorphisms specific for a particular individual.

The following examples are provided to illustrate certain aspects of the invention, and should not be construed in any way to limit the scope of the invention.

EXAMPLE 1

Genotyping of the Chr22.53 SNP, the CYP4.2D6.G4268C SNP

and the CYP450.2D6.G1749C SNP

I. Materials and Methods

A. Amplification of Target Nucleic Acid

Amplimers were generated through the PCR (Polymerase-Chain-Reaction) process using the HotStart Taq system from Qiagen. The PCR primer binding sites for the three SNPs analyzed, Chr. 22.53 SNP, CYP4.2D6.G4268C, and CYP450.2D6.G1749C SNP are shown in FIGS. 2, 5 and 7, respectively. Amplifications were conducted according to the manufacturer's recommendations. Primers were obtained from Operon Technologies, Inc. or synthesized using an ABI3948 Nucleic Acid Synthesis & Purification System. Typically, 40 cycles of PCR were performed to give a yield of 5-20 ng/μl DNA.

Excess dNTPs and primers were removed from the PCR products by the addition of exonuclease I and shrimp alkaline phosphatase (USB) at 1 unit/10 μl. After proper mixing, the tubes were incubated at 37 °C for 1 hour followed by 15 minutes at 80 C. At this stage, a small fraction of each sample was analyzed by agarose gel electrophoresis to determine the quality and quantity of the PCR products.

B. Primer Extension

For primer extension, approximately 10 ng of each PCR amplimer was used in a 5–μl reaction containing the following: 1x buffer for Taq FS as recommended by Perkin Elmer, 0.2 μM extension primer, 0.25 unit Taq FS (Perkin Elmer), and dye-labelled dideoxynucleotide triphosphates (NEN Life Science Products, Inc.) at the concentrations indicated in the figure legends. Extension was performed for 30 cycles in a GeneAmp PCR System 9700 (Perkin Elmer). Primer binding sites for the three SNPs studied are shown in FIGS. 2, 5 and 7.

C. Analysis of Extension Products

After desalting via ethanol precipitation, the samples were analyzed on a sequencing gel using the ABI 377 Sequencer. Results were plotted as fluorescence intensity versus time. The results shown (see FIGS. 3A, 3B, 4A, 4B, 6A, 6B, 8A and 8B) have been corrected for spectrum overlaps of the two dyes used.

II. Chr.22.53 Genotyping Results

A series of genotyping experiments for the Chr. 22.53 SNP (see FIG. 2) were conducted to compare the accuracy of genotyping results and the ability to control signal intensity with methods using substantially pure labelled forms of nucleotides and methods using mixtures of labelled and unlabelled forms of nucleotides. As used to refer to the purity of labelled nucleotides, the term "substantially pure" means that the unlabelled form of the nucleotide was only present at the trace levels found in commercially available products. Typically, unlabelled forms are approximately 5% of the total labelled and unlabelled nucleotides on a molar basis.

A. R110 and TAMRA Labeled Nucleotides

A first experiment was conducted using substantially pure labelled nucleotides. FIG. 3A shows electropherograms as plots of fluorescence intensity versus time for this set of genotyping experiments conducted with 7 different individuals at the Chr.22.53 SNP site. Experiments were conducted using 200 nM TAMRA-ddATP and 100 nM R110-ddGTP. Signals from extension products bearing the label R110 generated from the G allele are indicated by traces labelled G. Fluorescent signals for extension products bearing the label TAMRA generated from the A allele are shown by traces labelled A). The genotypes, as listed on the left side of each panel, were confirmed by sequencing the amplimers directly. As can be seen in this figure, the genotype was correctly identified for each of the 7 individuals. However, the signal intensities for the different alleles differ. Significant differences in signal intensities can complicate computerized analysis of signal traces. For example, in the G/A heterozygotes, the intensity of the A peak is 30-50% of the G peak. Small A peak is also observable in the G/G homozygote. An unsophisticated auto-allele-calling program may not be able to make accurate genotype determination in certain instances.

A second set of genotyping experiments were conducted with the same individuals and at the same SNP site but using a mixture of labelled and unlabelled nucleotides. More specifically, extension reactions were conducted in the presence of a mixture of 15 nM R110-ddGTP/85 nM ddGTP and a mixture of 30 nM TAMRA-ddATP/70 nM ddATP. Results showing extension products formed for different individuals is

summarized in the electropherograms shown in FIG. 3B. Extension product bearing the R110 label is indicated by traces labelled G and represents extension product formed from the G allele. Extension product bearing the TAMRA label is represented by traces labelled A and corresponds to the A allele. The correct genotypes are listed on the left side of the panels. Comparison of FIG. 3A and FIG. 3B indicate that equitable genotyping results are obtained in the presence of a mixture of labelled and unlabelled nucleotides. This comparison also demonstrates the ability to regulate the relative concentrations of the labelled nucleotides such that the signal intensities for different extension products is similar.

B. R110 and R6G Labeled Nucleotides

A third experiment with the same 7 individuals and SNP site was conducted using substantially pure labelled nucleotides. However, instead of using R110 and TAMRA labelled nucleotides as with the preceding experiments, R110 and R6G labelled nucleotides were utilized. Experiments were conducted using 100 nM R6G-ddATP and 100 nM R110-ddGTP. Electropherograms of the extension products formed for the 7 individuals is shown in FIG. 4A. Signals from extension products bearing the R6G label generated from the A allele are represented by traces labelled A. Extension products bearing the R110 label formed from the G allele are represented by traces labelled G. The correct genotypes, as confirmed by sequencing, are listed on the left of the panels. As can be seen in this figure, R110-ddGTP was misincorporated into individuals that were "AA" homozygotes.

A fourth genotyping experiment similar to the third experiment was conducted but with a mixture of labelled and unlabelled nucleotides. In particular, experiments were performed using a mixture of 15 nM R110-ddGTP/85 nM ddGTP and a mixture of 15 nM R6G-ddATP/85 nM ddATP. Results of the extension reactions are shown in FIG. 4B. Extension products labelled with R110 (corresponding to product from the G allele) are indicated by the traces labelled G. Extension products labelled with R6G (corresponding to product from the A allele) are represented by the traces labelled A. The correct genotypes are listed on the left of the traces. As a comparison of FIG. 4A and 4B illustrates, R110-ddG misincorporation into the "AA" homozygotes was significantly suppressed using mixtures of labelled and unlabelled nucleotides as compared to methods using substantially pure labelled nucleotides.

III. CYP4.2D6.G4268C Genotyping Results

Another set of genotyping experiments were conducted with 7 individuals at the CYP4.2D6.G4268C SNP (Sachse *et al.*, *Am. J. Hum. Genet.* 6:284-295 (1997), which is incorporated by reference in its entirety). The sequence surrounding this SNP is shown in FIG. 5. This experiment investigated the ability to utilize mixtures of labelled and unlabelled forms of a deoxynucleotide in combination with a dideoxynucleotide.

A. Combination of Dideoxynucleotides

One experiment was conducted using a mixture of labelled and unlabelled nucleotides, specifically 10 nM R110-ddCTP/ 90 nM ddCTP and a mixture of 30 nM TAMRA-ddGTP/ 70 nM ddGTP. FIG. 6A presents the electropherograms obtained for the different individuals. Signals from extension products incorporating R110 label and corresponding to the C allele are represented by the traces labelled with a C. Extension products labelled with TAMRA and corresponding to the G allele are represented by traces labelled G. The genotypes, as indicated in parentheses, were also confirmed by sequencing the amplimers directly.

B. Combination of Dideoxynucleotide and Deoxynucleotide

Another experiment was conducted using mixtures of labelled and unlabelled forms of a dideoxynucleotide and a deoxynucleotide. Experiments were conducted with a mixture of 10 nM R110-dCTP/90 nM dCTP and mixtures of 30 nM TAMRA-ddGTP/70 nM ddGTP. FIG. 6B depicts the electropherograms obtained for each of the 7 individuals tested. Extension products incorporating R110 and generated from the C allele are represented by traces labelled C. Extension products labelled with TAMRA and generated from the G allele are represented by traces labelled G. The correct genotypes are as indicated in parentheses. A comparison of FIG. 6A and 6B demonstrates that a deoxynucleotide can also be used in the primer extension reaction to obtain the correct genotypes.

IV. CYP450.2D6.G1749C SNP Genotyping Results

A. Labelled Nucleotides

One set of genotyping experiments was conducted with 7 individuals at the CYP450.2D6.G1749C SNP (Sachse *et al.*, *Am. J. Hum. Genet.* 6:284-295 (1997), which is

incorporated by reference in its entirety). The sequence surrounding this SNP is shown in FIG. 7.

A first experiment was conducted using substantially pure labelled nucleotides. FIG. 8A shows electropherograms as plots of fluorescence intensity versus time for this set of genotyping experiments conducted with 7 different individuals at the CYP450.2D6.G1749C site. Experiments were conducted using 100 nM R110-ddCTP and 100 nM R6G-ddGTP. Signals from extension products bearing the label R110 generated from the C allele are indicated by traces labelled C. Fluorescent signals for extension products bearing the label R6G generated from the G allele are shown by traces labelled G. The genotypes, as listed on the left side of each panel, were confirmed by sequencing the amplimers directly. As can be seen in FIG. 8A, the genotype was correctly identified for each of the 7 individuals.

B. Mixture of Labelled and Unlabelled Nucleotides

A second set of genotyping experiments was conducted with the same individuals and at the CYP4.2D6.G1749C SNP site but using a mixture of labelled and unlabelled nucleotides. More specifically, extension reactions were conducted in the presence of a mixture of 10 nM R110-ddCTP/90 nM ddCTP and a mixture of 10 nM TAMRA-ddGTP/90 nM ddGTP. Results showing extension products formed for different individuals is summarized in the electropherograms shown in FIG. 8B. Extension product bearing the R110 label is indicated by traces labelled C and represents extension product formed from the C allele. Extension product bearing the R6G label is represented by traces labelled G and corresponds to the G allele. The correct genotypes are listed on the left side of the panels.

A comparison of FIG. 8A and FIG. 8B demonstrates that the results obtained with a mixture of labeled and unlabeled nucleotides is consistent with the results obtained using just labelled nucleotides.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents and patent

applications cited herein are hereby incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent or patent application were specifically and individually indicated to be so incorporated by reference.

WHAT IS CLAIMED IS:

1. A method of analyzing a site of variation in a target polynucleotide, comprising:

(a) contacting a sample of the polynucleotide with multiple copies of a primer that hybridizes adjacent to but not including the site of variation in the presence of a mixture of labelled and unlabelled forms of a nucleotide, wherein the unlabelled form is 20% to 95% of the total amount of the labelled and unlabelled forms on a molar basis, the contacting being performed under conditions whereby if the nucleotide is complementary to the base occupying the site of variation in the target polynucleotide, at least one copy of the primer is extended by incorporation of the labelled nucleotide; and

(b) detecting incorporation of the labeled nucleotide into the primer as an indication of the base occupying the site of variation in the target polynucleotide.

2. The method of claim 1, wherein the labelled and unlabelled forms of a nucleotide are labelled and unlabelled forms of a non-extendible nucleotide.

3. The method of claim 1, wherein the non-extendible nucleotide is a dideoxynucleotide or an arabinoside triphosphate.

4. The method of claim 1, wherein the labelled and unlabelled forms of a nucleotide are labelled and unlabelled forms of an extendible nucleotide.

5. The method of claim 1, wherein the unlabelled form is 50% to 90% of the total amount of labelled and unlabelled forms on a molar basis.

6. The method of claim 1, further comprising mixing unlabelled and labelled forms of nucleotide from separate sources.

7. The method of claim 1, wherein the conditions comprise presence of a thermostable polymerase, and performing multiple cycles of denaturation, annealing and primer extension.

8. The method of claim 7, wherein the thermostable polymerase is selected from the group consisting of Taq, Taq FS and ThermoSequenase.

9. The method of claim 1, wherein the conditions comprise presence of a non-thermostable polymerase and performing a single cycle of primer extension.

10. The method of claim 9, wherein the non-thermostable polymerase is selected from the group consisting of Klenow fragment, Bst Polymerase (large fragment) and Sequenase.

11. The method of claim 1, wherein the label is selected from the group consisting of a fluorophore, a chromophore, a radioisotope, an enzyme substrate, an electron dense agent, a magnetic particle, an electrochemically active moiety and a mass label.

12. The method according to claim 11, wherein the label is a fluorescent dye.

13. The method according to claim 12, wherein the fluorescent dye is selected from the group consisting of FAM, ROX, TAMRA, R110, R6G, Joe, TET, HEX, Alexa, Cy3 and Cy 5.

14. The method of claim 1, wherein the multiple copies of primer are labeled with a primer label different from the label borne by the labelled nucleotide, whereby if the nucleotide is complementary to the base occupying the site of variation, then at least one copy of extended primer bears labels from both labelled primer and labelled nucleotide.

15. The method of claim 14, wherein the primer label and the label borne by the labelled nucleotide comprise a donor fluorophore and an acceptor fluorophore.

16. The method of claim 1, wherein detection of labelled nucleotide incorporation is performed by fluorescence polarization.

17. A method of analyzing a site of variation in a target polynucleotide, comprising:

(a) contacting a sample of the polynucleotide with multiple copies of a primer that hybridizes adjacent to but not including the site of variation in the presence of a plurality of mixtures of labelled and unlabelled forms of a nucleotide, each mixture

containing a different nucleotide, for each nucleotide the unlabelled form being 20% to 95% of the combined amount of the labelled and unlabelled forms on a molar basis, and the contacting step being performed under conditions whereby if the nucleotide of a mixture is complementary to the base occupying the site of variation in the target polynucleotide, at least one copy of the primer is extended by incorporation of the labelled nucleotide; and

(b) detecting the incorporation of labelled nucleotide into the primer(s) extended by the labelled nucleotide(s) as an indication of the base(s) occupying the site of variation in the target polynucleotide.

18. The method of claim 17, wherein each of the plurality of mixtures comprises a mixture of labelled and unlabelled non-extendible nucleotides.

19. The method of claim 17, wherein each of the plurality of mixtures comprises a mixture of labelled and unlabelled extendible nucleotides.

20. The method of claim 18, wherein the non-extendible nucleotides are selected from the group consisting of a dideoxynucleotide, and an arabinoside triphosphate.

21. The method of claim 17, wherein for each of the plurality of mixtures the unlabelled form is 50% to 90% of the total amount of labelled and unlabelled forms on a molar basis.

22. The method of claim 17, wherein the plurality of mixtures is 2, 3 or 4 mixtures.

23. The method of claim 17, wherein the contacting step comprises dividing the sample into a plurality of separate samples such that the number of samples is equivalent to the number of mixtures of nucleotides and contacting each of the samples with a different mixture of nucleotides.

24. The method of claim 17, wherein the plurality of mixtures is a first mixture of labelled and unlabelled forms of a first nucleotide and a second mixture of labelled and unlabelled forms of a second nucleotide.

25. The method of 24, wherein the molar ratio of unlabelled to labelled forms in the first mixture is at least 1:2 and the ratio of the unlabelled to labelled forms in the second mixture is at least 1:2.

26. The method of claim 17, wherein the sample of the target polynucleotide is a heterozygous sample comprising first and second target polynucleotides differing at the site of variation, and at least one copy of primer hybridizing to the first target polynucleotide is extended by incorporation of the first labelled nucleotide, and at least one copy of primer hybridizing to the second target polynucleotide is extended by incorporation of the second labelled nucleotide.

27. The method of claim 17, wherein the target polynucleotide is from a diploid subject and comprises a first and/or second target polynucleotide that differ at the site of variation, whereby if the first target polynucleotide is present, at least one copy of primer hybridizes to the first target polynucleotide and is extended by incorporation of the first labelled nucleotide, and if the second target polynucleotide is present, at least one copy of primer hybridizes to the second target polynucleotide and is extended by incorporation of the second labelled nucleotide.

28. The method of claim 27, wherein the detecting step comprises detecting the presence of primers extended by the first and/or second labelled nucleotide, the presence of primers extended only by the incorporation of first labelled nucleotide or second labelled nucleotide indicating that the subject is a homozygote, whereas the presence of primers extended by the incorporation of first labelled nucleotide and primers extended by second labelled nucleotide indicates that the subject is a heterozygote.

29. The method of claim 24, wherein the detecting step comprises detecting a first signal from primer extended by the first labelled nucleotide and a second signal from primer extended by the second labelled nucleotide, and the molar ratio of the first and second labeled nucleotides is such that the first and second signals are of similar strength.

30. The method of claim 17, wherein the conditions include presence of a thermostable polymerase, and performing multiple cycles of denaturation, annealing and primer extension.

31. In a method of analyzing a site of variation in a target polynucleotide comprising contacting the target polynucleotide with a primer hybridizing adjacent to but not including the site of variation and a labelled nucleotide under conditions in which the primer is extended by the labelled nucleotide if the labelled nucleotide is complementary to the site of variation, the improvement wherein the target polynucleotide is contacted with a mixture of labelled and unlabelled forms of the nucleotide and the unlabelled form is at least 30% of the combined amount of the labelled and unlabelled forms on a molar basis.

32. A method for analyzing sites of variation in target polynucleotides, comprising:

(a) conducting a plurality of template-dependent extension reactions with a plurality of different primers and one or more mixtures of labelled and unlabelled forms of a nucleotide, wherein different primers hybridize adjacent different sites of variation on target polynucleotides and different mixtures contain different nucleotides, the unlabelled form in each mixture being 20% to 90% of the combined amount of the labelled and unlabelled forms on a molar basis, each extension reaction comprising

(i) contacting a sample of the target polynucleotides with multiple copies of one of the different primers, wherein the primer bears an optional label and the 3' end of the primer hybridizes adjacent to but not including the site of variation in the target polynucleotide, and

(ii) exposing the copies of the primer to one or more of the mixtures of labelled and unlabelled nucleotides, under conditions such that if the nucleotide of the mixture is complementary to the base occupying the site of variation in the target polynucleotide, at least one copy of the primer is extended by incorporation of the labelled nucleotide,

whereby the extension reactions generate a plurality of different extension products, extension products generated from different sites of variation being distinguishable

on the basis of the optional label borne by the extended primer and/or label borne by incorporated nucleotide; and

(b) detecting the incorporation of labelled nucleotides into the extended primers as an indication of the bases occupying the site of variation in the target polynucleotides.

33. The method of claim 32, wherein the optional label is selected from the group consisting of a fluorophore, a chromophore, a radioisotope, an enzyme substrate, an electron dense agent, a magnetic particle, an electrochemically active moiety and a mass label.

34. The method of claim 32, wherein the different sites of variation are different sites on the same target polynucleotide or same site on different polynucleotides and the extension reactions are conducted in a single reaction vessel.

35. The method of claim 32, further comprising separating the different extension products prior to detecting incorporation of labelled nucleotide, separation being accomplished by electrophoresis or high pressure liquid chromatography.

36. The method of claim 32, wherein

(i) the multiple copies of primer are labeled with a primer label, whereby extension products bear two labels, a primer label and label from incorporated labelled nucleotide, the two labels comprising a donor fluorophore and an acceptor fluorophore; and

(ii) detecting comprises detecting changes in fluorescence of the donor and/or acceptor fluorophores.

37. The method of claim 32, wherein detection of labelled nucleotide incorporation is performed by fluorescence polarization.

38. The method of claim 32, wherein the different sites of variation are located on different target polynucleotides and each extension reaction is conducted in a separate reaction vessel, and further comprising collecting extension product from the reaction vessels prior to detection.

39. A composition comprising a mixture of labelled and unlabelled forms of a nucleotide, wherein the unlabelled form is 30% to 95% of the combined amount of labelled and unlabelled forms on a molar basis.

40. The composition of claim 39, wherein the unlabelled nucleotide is a non-extendible nucleotide.

41. The composition of claim 39, wherein the unlabelled nucleotide is an extendible nucleotide.

42. The composition of claim 39, wherein the label borne by the labelled nucleotide is a fluorophore.

43. The composition of claim 42, wherein the fluorophore is selected from the group consisting of FAM, ROX, TAMRA, R110, R6G, Joe, HEX, TET, Alexa, Cy3 and Cy 5.

44. A kit for analyzing a variant site in a target nucleic acid, comprising:

(a) a mixture of labelled and unlabelled nucleotide, the unlabelled nucleotide being 30% to 95% of the combined amount of unlabelled and labelled forms of the nucleotide on a molar basis;

(b) a polymerase capable of catalyzing template-dependent extension of a primer.

(c) a primer that hybridizes to a segment of the target nucleic acid, wherein the 3' end of the primer hybridizes adjacent to the site of variation in the target nucleic acid and bears an optional label.

45. The kit of claim 44, wherein the label borne by the labelled nucleotide is a fluorophore.

46. The kit of claim 45, wherein the fluorophore is selected from the group consisting of FAM, ROX, TAMRA, R110, R6G, Joe, HEX, TET, Alexa, Cy3 and Cy 5.

47. The kit of claim 44, wherein the polymerase is a thermostable polymerase.

48. The kit of claim 44, wherein the 3' end of the primer hybridizes immediately adjacent to the site of variation in the target nucleic acid.

49. The kit of claim 48, wherein the primer is at least 5 primers, each primer hybridizing adjacent to a different variant site in the target nucleic acid.

50. The kit of claim 49, wherein each variant site is a polymorphic site correlated with a disease.

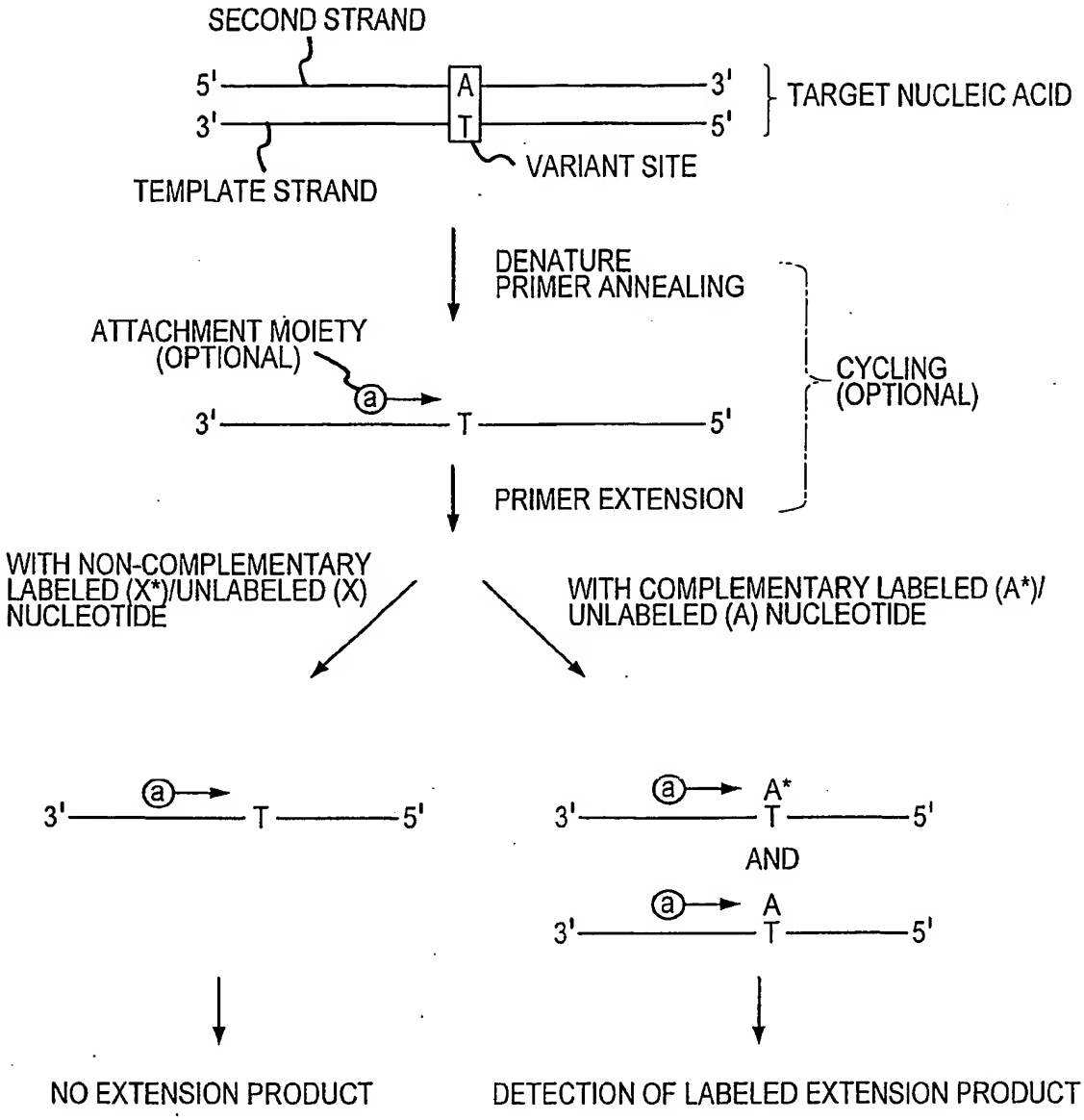

SECOND STRAND
5'
3'
A
T
3'
5'
TARGET NUCLEIC ACID
TEMPLATE STRAND
VARIANT SITE
DENATURE
PRIMER ANNEALING
ATTACHMENT MOIETY
(OPTIONAL)
a
3'
T
5'
CYCLING
(OPTIONAL)
PRIMER EXTENSION
WITH NON-COMPLEMENTARY
LABELED (X*)/UNLABELED (X)
NUCLEOTIDE
WITH COMPLEMENTARY LABELED (A*)/
UNLABELED (A) NUCLEOTIDE
a
3'
T
5'
a
A*
3'
T
5'
AND
a
A
3'
T
5'
NO EXTENSION PRODUCT
DETECTION OF LABELED EXTENSION PRODUCT

FIG.1

aaggcaagtg acctggcaaa ggcctcccag catatatggt cacaggtaaa
actcaagcct ggtctgttat tccctcccca aaactgcctg gtagggagaa
gggaaagcca gcctc (R) gctgtgactg actcctggct atctggagct
ctccctccag attcctttct acctctctgg ctgtcacccc agtgctaaga
gcctgggctc aggccctcct ctctctaccc tggctgctcc cgctcattcc
tgtggcttct gtccccagta ctctagca

FIG.2

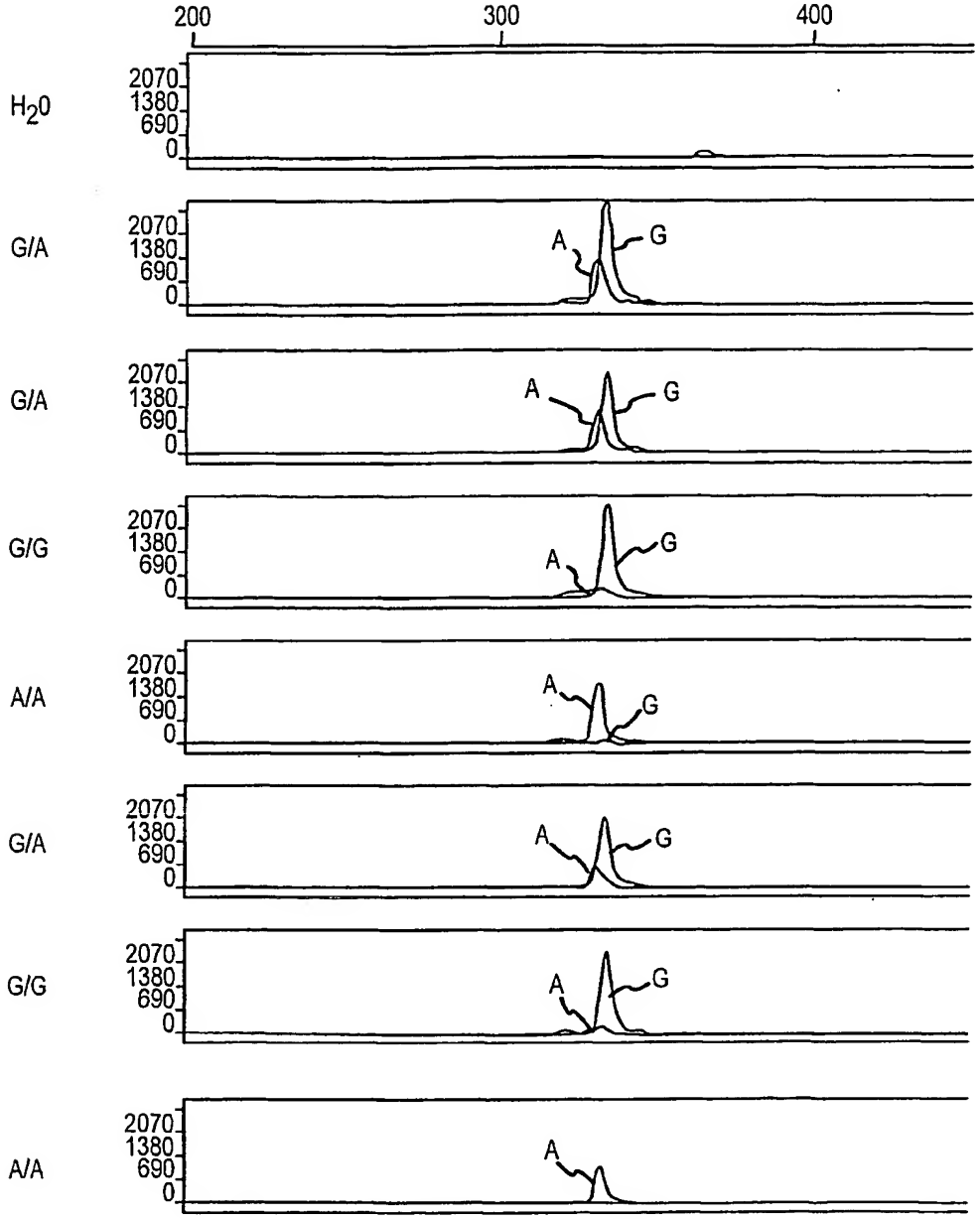


FIG.3A

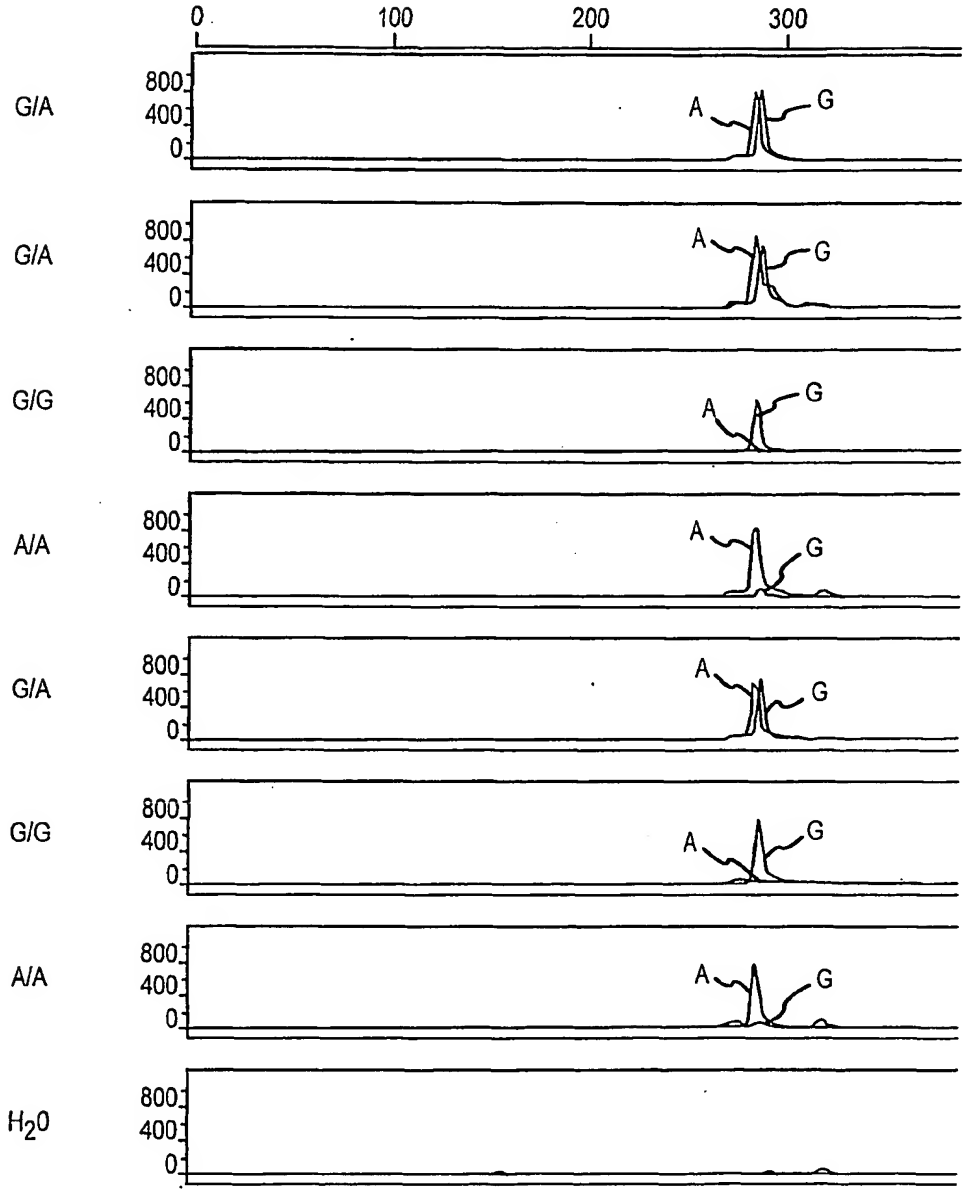


FIG.3B

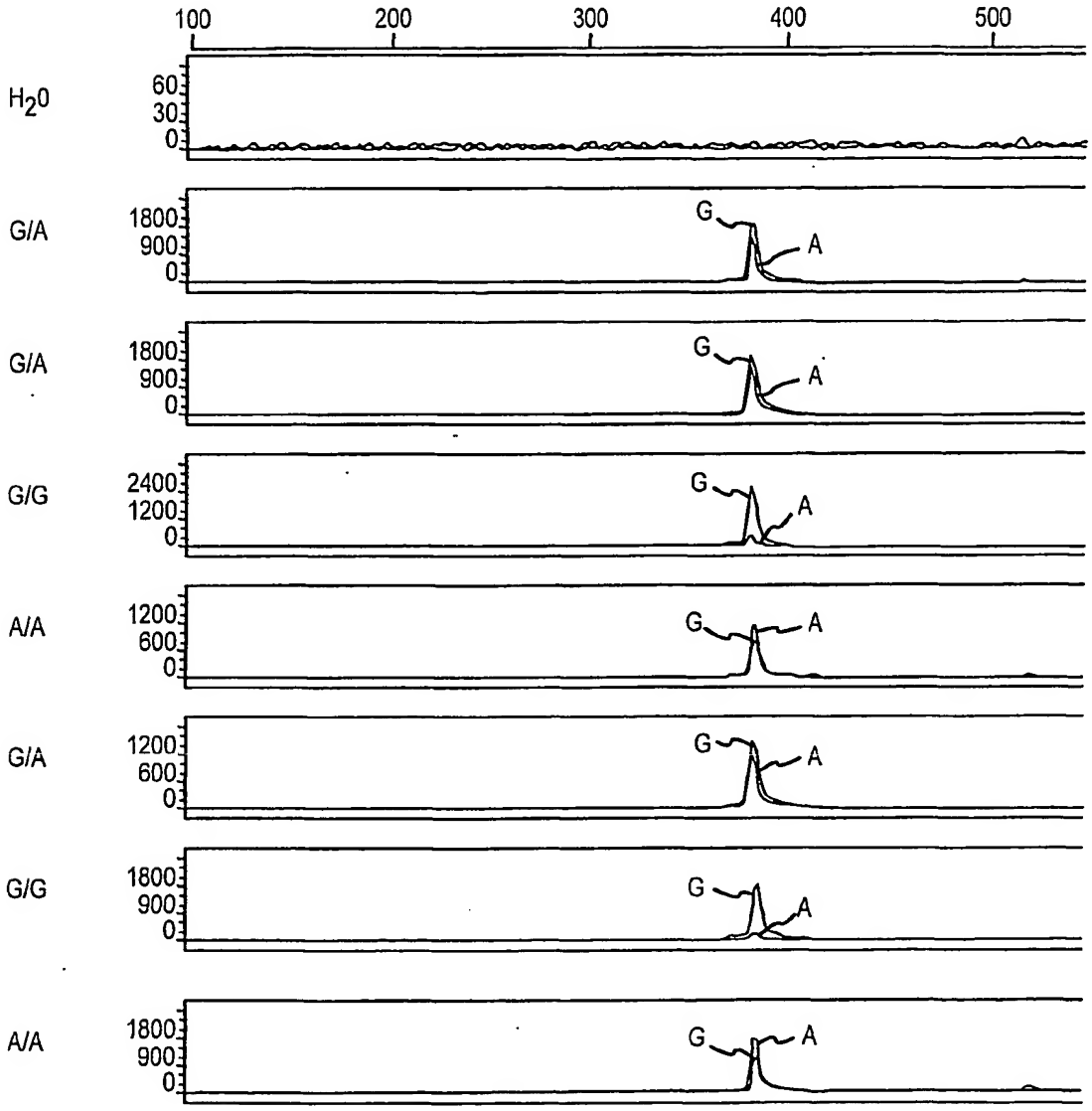

100
200
300
400
500
H20
60
30
0
G/A
1800
900
0
G
A
G/A
1800
900
0
G
A
G/G
2400
1200
0
G
A
A/A
1200
600
0
G
A
G/A
1200
600
0
G
A
G/G
1800
900
0
G
A
A/A
1800
900
0
G
A

FIG.4A

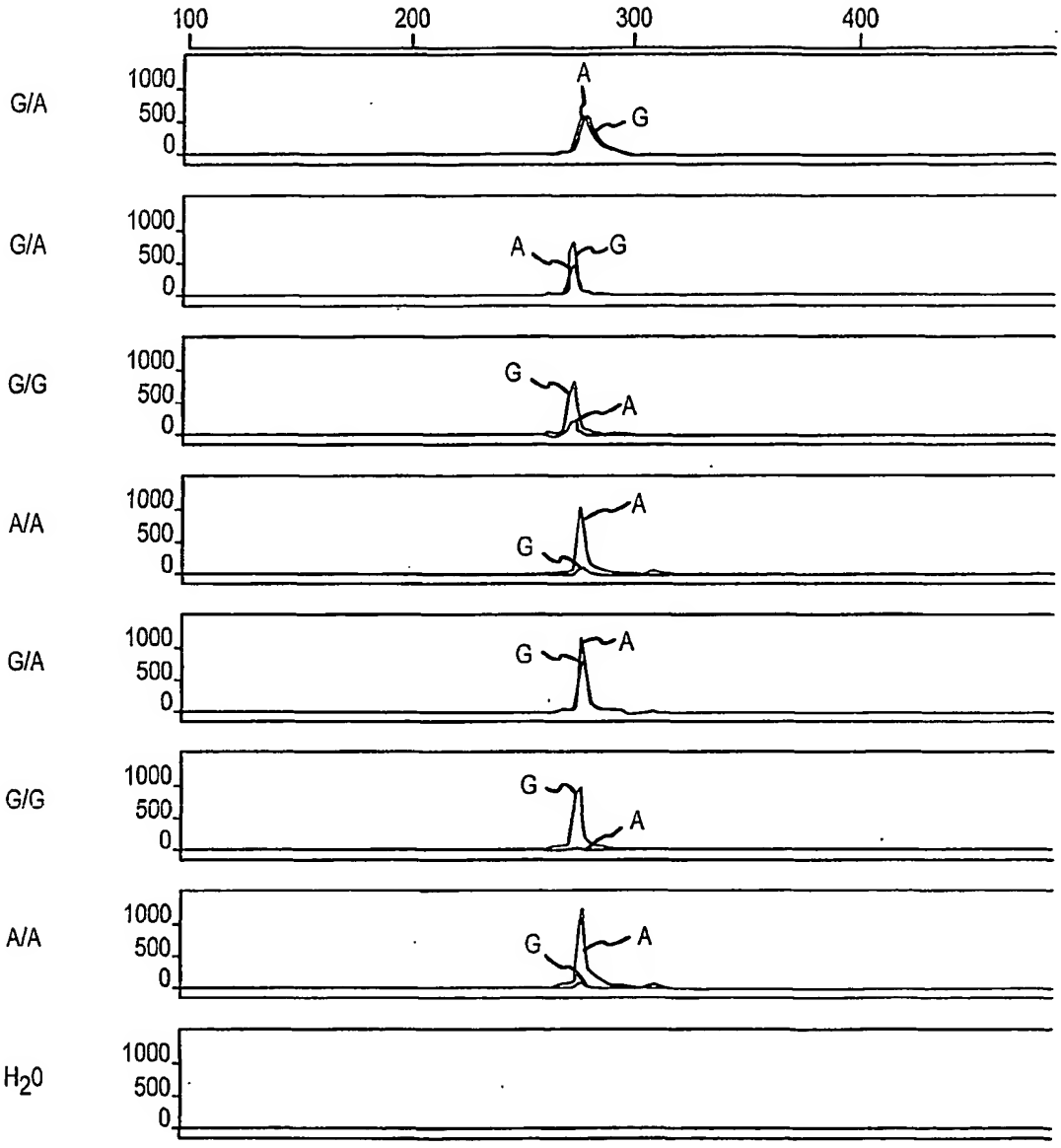

100
200
300
400
G/A
1000
500
0
A
G
G/A
A
G
G/G
G
A
A/A
G
A
G/A
G
A
G/G
G
A
A/A
G
A
H_2O

FIG.4B

5′ cttccgtgga gtcttgcagg ggtatcaccc aggagccagg
ctcactgacg cccctcccct ccccacaggc cgccgtgcat
gcctcgggga gcccctggcc cgcatggagc tcttcctctt
cttcacctcc ctgctgcagc acttcagctt ctcggtgccc
actggacagc cccggcccag ccaccatggt gtctttgctt
tcctggtgaS cccatccccc tatgagcttt gtgctgtgcc
ccgctagaat ggggtaccta gtccccagcc tgctccctag
ccagaggctc taatgtacaa taaagcaatg tggtagttcc
aactcgggtc ccctgctcac gccctcgttg ggatcatcct
cctcagggca accccacccc tgcctcattc ctgcttaccc
caccgcctgg ccgcatttga gacaggggta cgttgaggct
gag 3′

FIG.5

8/12

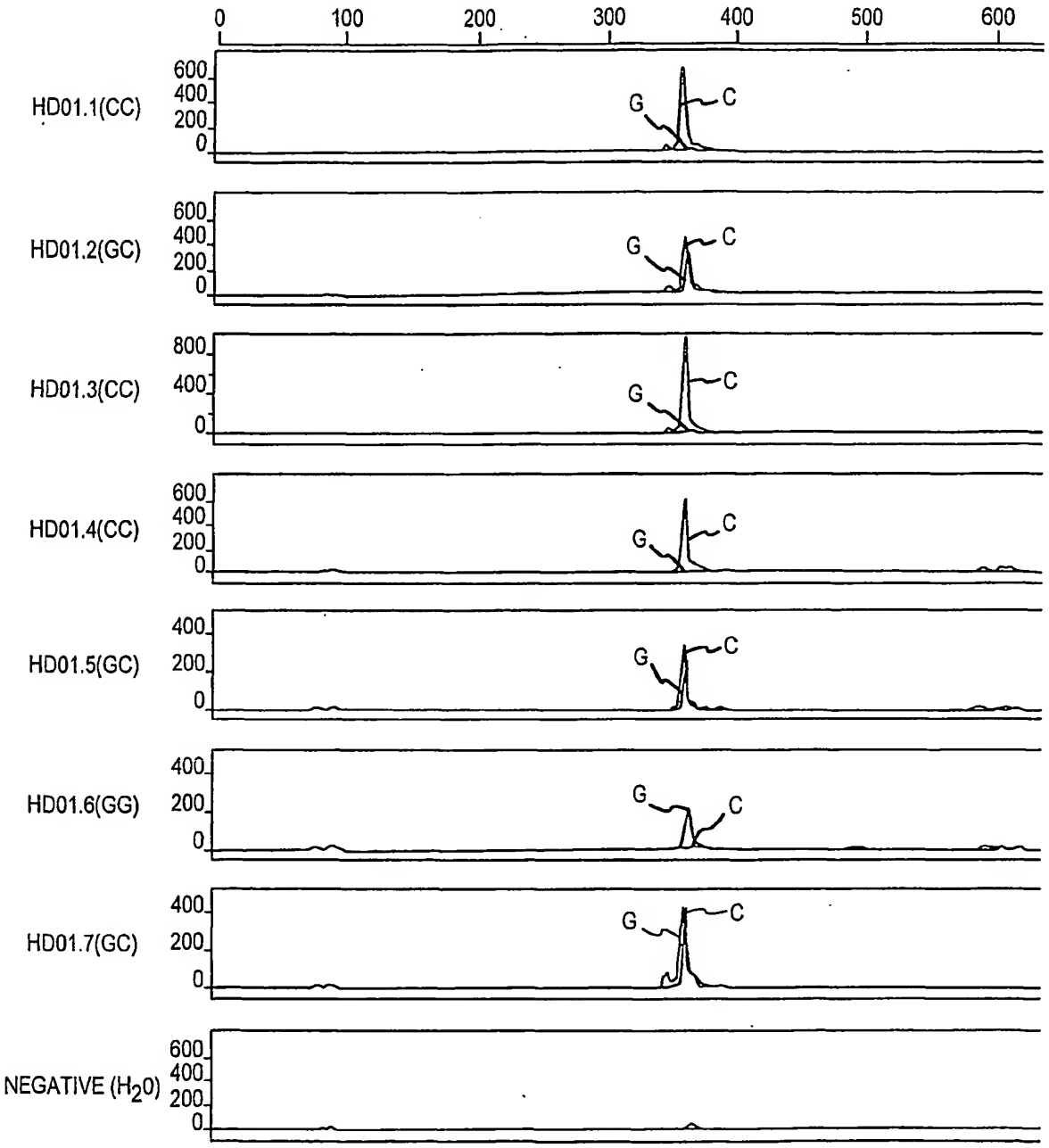

0
100
200
300
400
500
600
HD01.1(CC)
HD01.2(GC)
HD01.3(CC)
HD01.4(CC)
HD01.5(GC)
HD01.6(GG)
HD01.7(GC)
NEGATIVE (H_2O)
G
C

FIG.6A

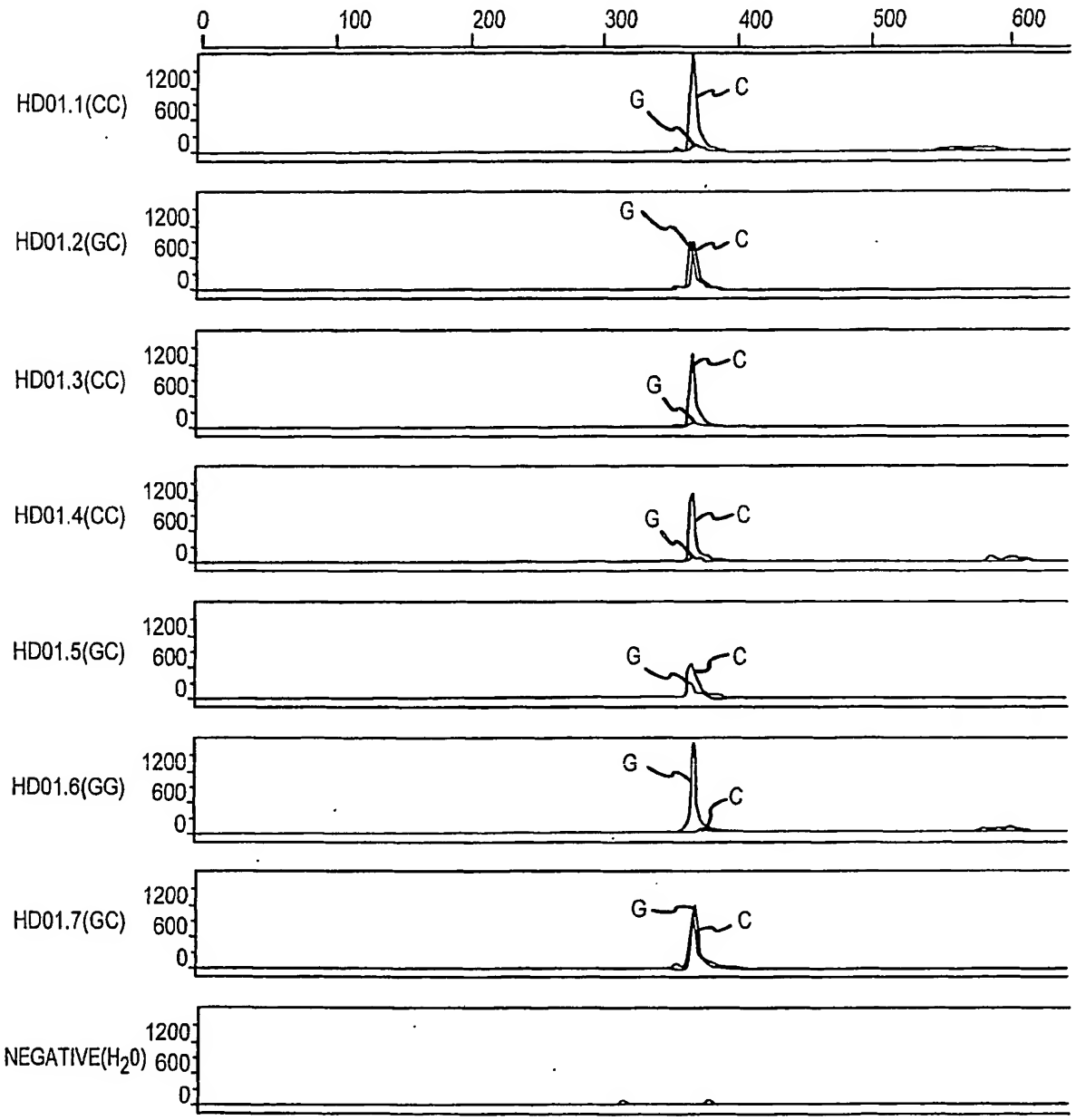

0
100
200
300
400
500
600
HD01.1(CC)
HD01.2(GC)
HD01.3(CC)
HD01.4(CC)
HD01.5(GC)
HD01.6(GG)
HD01.7(GC)
NEGATIVE(H_2O)
1200
600
0
G
C

FIG.6B

agaccagggg gagcataggg ttggagtggg tggt**ggatgg tggggctaat**
gccttcatgg ccacgcgcac gtgcccgtcc caccccccagg ggtgttcctg
gcgcgctatg ggcccgcgtg gcgcgagcag aggcgcttct ccgt**S**tccac
cttgcgcaac ttgggcctgg gcaagaagtc gctggagcag ngggtgaccg
aggaggccgc ctgcctttgt gccgccttcg ccaaccactc ckgtgggtga
tgggcagaag ggcacaaagc gggaactggg aaggcggggg acggggaagg
cgacccctta cccgcatctc ccacccccar gacgcccctt tcgc**ccccaac**
ggtctcttgg acaaagccgt gagcaacgtg atcgcctccc tcacctgcgg

FIG.7

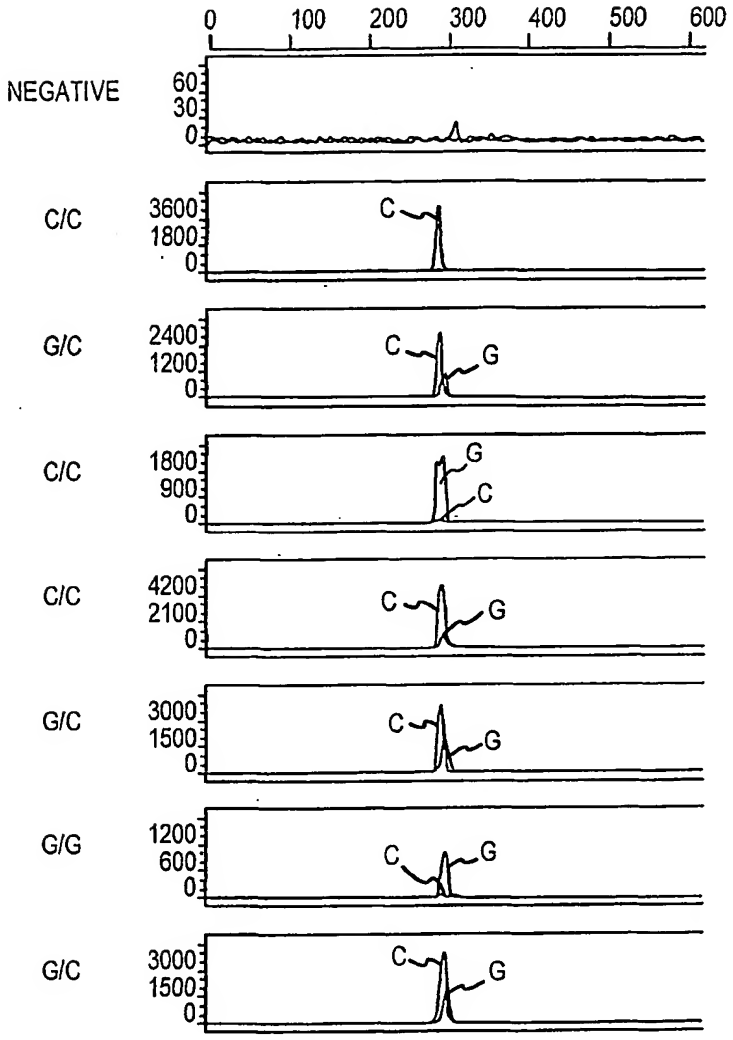

0
100
200
300
400
500
600
NEGATIVE
60
30
0
C/C
3600
1800
0
C
G/C
2400
1200
0
C
G
C/C
1800
900
0
G
C
C/C
4200
2100
0
C
G
G/C
3000
1500
0
C
G
G/G
1200
600
0
C
G
G/C
3000
1500
0
C
G

FIG.8A

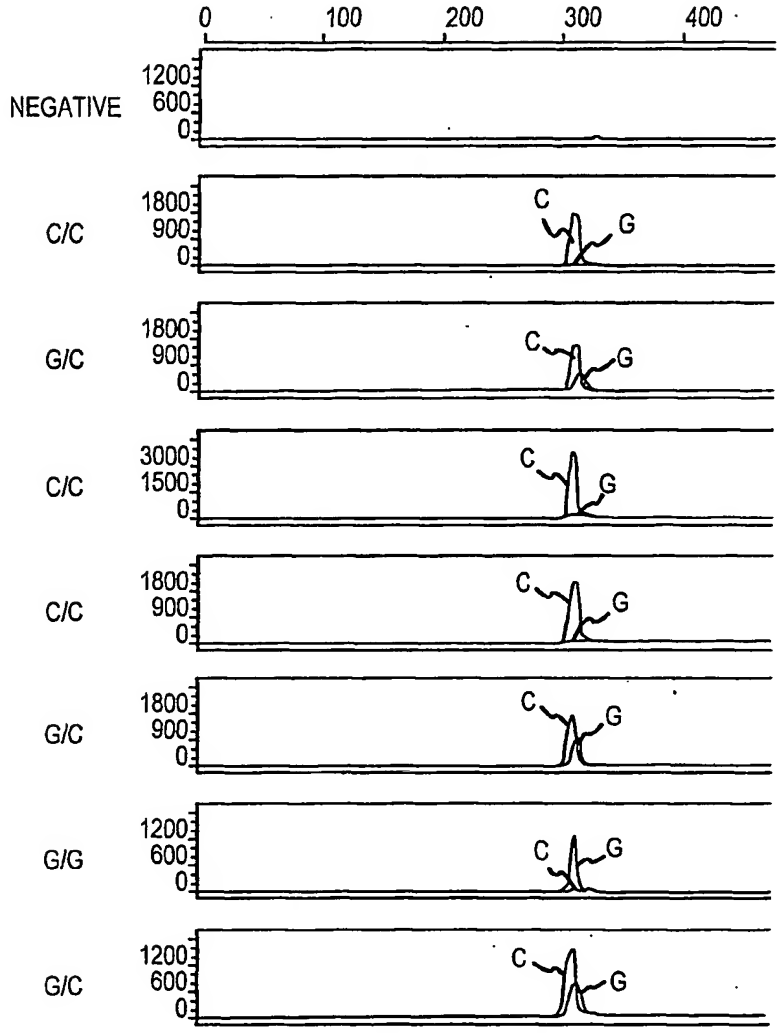

0
100
200
300
400
NEGATIVE
1200
600
0
C/C
1800
900
0
C
G
G/C
1800
900
0
C
G
C/C
3000
1500
0
C
G
C/C
1800
900
0
C
G
G/C
1800
900
0
C
G
G/G
1200
600
0
C
G
G/C
1200
600
0
C
G

FIG.8B